

Current insights into code-switching

Edited by

Draško Kaščelan

María del Carmen Parafita Couto

Peredur Webb-Davies

Rocío Pérez-Tattam

Hans Stadthagen-González

Current Issues in Bilingualism ??

Current Issues in Bilingualism

Editors: Andrea C. Schalley (Karlstad University, Sweden) [editor-in-chief],
M Carmen Parafita Couto (Leiden University, Netherlands) [editor-in-chief],
Susana Eisenclas (Griffith University, Australia),
Galina Putjata (Europa-Universität Flensburg, Germany),
Jorge Valdés Kroff (University of Florida, USA).

In this series:

1. Shim, Ji Young. OV and VO variation in code-switching.
2. Rodríguez González, Eva & Rosita L. Rivera (eds.). Language assessment in multilingual settings: Innovative practices across formal and informal environments.

ISSN (print): 2747-9919

ISSN (electronic): 2747-9927

Current insights into code-switching

Edited by

Draško Kaščelan

María del Carmen Parafita Couto

Peredur Webb-Davies

Rocío Pérez-Tattam

Hans Stadthagen-González

Draško Kaščelan, María del Carmen Parafita Couto, Peredur Webb-Davies, Rocío Pérez-Tattam & Hans Stadthagen-González (eds.). 2025. *Current insights into code-switching* (Current Issues in Bilingualism ??). Berlin: Language Science Press.

This title can be downloaded at:

`redefine\lsURL`

© 2025, the authors

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/>

ISBN: `no digital ISBN`

`no print ISBNs!`

ISSN (print): 2747-9919

ISSN (electronic): 2747-9927

`no DOI`

Cover and concept of design: Ulrike Harbort

Fonts: Libertinus, Arimo, DejaVu Sans Mono

Typesetting software: Xe_{La}TeX

`redefine \publisherstreetaddress`

`redefine \publisherurl`

Storage and cataloguing done by `redefine \storageinstitution`

Contents

Acknowledgments	iii
1 Code-switching in context: Variation, processing, and development Peredur Webb-Davies, María del Carmen Parafita Couto, Draško Kaščelan	1
2 Code-mixing and language change: Evidence from Spanish nouns in heritage Piedmontese Eugenio Gorla & Kate Bellamy	17
3 <i>Verinaiss</i> ‘very nice’: The integration of English in Estonian teenagers’ instant messages Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli	45
4 Cued gender assignment strategies and L1 effects in the acceptance of mixed Basque-Spanish DPs Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola	89
5 Sensitivity to codeswitching asymmetries in second language processing Jorge R. Valdés Kroff, Hannah G. Treadway, Rosa E. Guzzardo Tamargo & Paola E. Dussias	127
6 Linguistic interaction in early bilingual development: A review of studies examining single- and mixed-language utterances Marilyn M. Vihman & Virve-Anneli Vihman	161
7 Supporting bilingual development through code-switching: Dual language use of a grandmother and an emergent bilingual child Suzanne Quay	209

Contents

**8 Code-switching by individuals with neurodevelopmental conditions:
A scoping review**
Draško Kaščelan & María del Carmen Parafita Couto **247**

Index **289**

Acknowledgments

Chapter 1

Code-switching in context: Variation, processing, and development

Peredur Webb-Davies, María del Carmen Parafita Couto,
Draško Kaščelan

Bangor University

1 Volume introduction

This volume provides an exploration of code-switching in bilinguals/multilinguals, examining the phenomenon in linguistic development, social interaction, and processing. The volume focuses on grammatical variation in speech language patterns, as well as the development of code-switching, including its occurrence in neurodevelopmental conditions. By integrating research across diverse bilingual contexts, this collection offers a comprehensive understanding of how code-switching functions within different linguistic communities.

Code-switching (or code-mixing) is a well-documented feature of the speech of multilinguals whereby words or grammar from multiple languages are combined in a single discourse, conversation or utterance. Use of code-switching is shaped by a confluence of factors, including cognitive development, linguistic structures and, crucially, the socio-cultural contexts in which bilinguals live and interact. The chapters in this volume investigate code-switching from different perspectives and explore other governing factors at play.

In grammatical terms, code-switching can be classified in various ways, perhaps the most salient being the distinction between inter-clausal switching (where the switch occurs at clause boundaries) and intra-clausal switching

Peredur Webb-Davies, María del Carmen Parafita Couto, Draško Kaščelan

(where the switch happens within a clause).¹ Research on code-switching has been prevalent in the literature for many decades, and since the 1980s a particular focus has been given to grammatical constraints (if any) on code-switching. A brief reference to the journey of this debate is given by Valdés Kroff et al. in chapter 5 of this volume. It will suffice to note at present that studies in this field have investigated the patterns of code-switching found in many diverse speech populations, with a particular focus since the work of Poplack (1980) on Spanish-English, but also a significant contribution to the study of the speech of Welsh-English bilinguals by Deuchar and colleagues (e.g., Deuchar et al. 2014, 2018). Generally, studies show that code-switching is not the same for all bilinguals (cf. the in-depth meta-review and typology provided by Muysken 2000), though it is by no means random and is in fact governed by factors such as community norms and the mapping of the grammars of the languages involved.

Chapters 2–8 in this volume comprise a range of studies examining code-switching from multiple perspectives. Key themes explored across the presented studies include:

- *Community norms and language practices:* Code-switching reflects the linguistic norms of a given community. When bilingualism is widely accepted and practiced, children are more likely to integrate code-switching into their speech. Conversely, in communities where strict language separation is emphasized, code-switching may be less frequent or carry a different social meaning (chapter 6). Code-switching for heritage communities (such as the Piedmontese-Spanish speaking communities in Argentina) can be limited to certain kinds of activities where the heritage language is prioritised, though switching still occurs (chapter 2). From a methodological standpoint, tasks that are minimally metalinguistic - such as a participants indicating they comprehend a stimulus - appear to be more in-line with the community speech norms of bilinguals, whereas more metalinguistic tasks provide different results (chapter 5).
- *Caregiver influence on bilingual development:* Parents and caregivers play a pivotal role in shaping children's code-switching behaviors. Their choices in providing bilingual input, whether through inter-sentential or intra-sentential switching, impact children's language development. Studies sug-

¹Note that across literature, there is a variation in whether a clause or a sentence is used as a unit of analysis. In case a sentence is used as a unit of analysis, intra-sentential switch would refer to a switch within a sentence (which could occur between clauses in case of complex or compound sentences), while inter-sentential switch would refer to a switch between sentences.

1 *Code-switching in context: Variation, processing, and development*

gest that children tend to accommodate their caregivers in speech practices, including the use of code-switching (chapter 7).

- *Attitudes and ideologies surrounding bilingualism:* Perceptions of code-switching vary widely. In some contexts, bilingualism is celebrated, and code-switching is viewed as a resourceful linguistic tool. In others, it is stigmatized as a sign of incomplete language acquisition. These attitudes shape language policies at family, school, and community levels, affecting bilingual children's linguistic experiences (chapter 8). Bilinguals recognize that code-switching within phrases (e.g., determiner phrases) occur in their community's speech, but they are found in a study in this volume to judge them less grammatically acceptable than unilingual utterances from one of their languages (chapter 4).
- *Code-switching as a communicative resource:* Rather than being a linguistic deficit, code-switching serves as a strategic tool for meaning-making and social identity negotiation. Bilingual children use code-switching to enhance their expressive capabilities and engage more effectively in communication (chapter 7). Young bilinguals who use instant messaging deploy code-switching creatively in terms of both word-selection, orthography and degree of morphological integration of switched items (chapter 3).
- *Neurodiversity and bilingualism:* By reviewing research on bilinguals with neurodevelopmental conditions, this volume broadens the understanding of code-switching. It challenges deficit-based perspectives that pathologize language mixing and highlights the importance of evidence-based guidance for multilingual families with neurodivergent family members (chapter 8).

The remainder of this chapter summarizes and contextualizes the studies presented in this book.

1.1 Grammatical variation and code-switching

The range of approaches to researching code-switching variation in the field is well-represented in chapters 2-5, with brief commentaries on those chapters provided here.

The study by Gorla and Bellamy (chapter 2) investigates code-mixing in the speech of Heritage Piedmontese-Argentinian Spanish bilinguals living in Argentina, specifically focusing on Spanish noun insertions into Piedmontese ma-

Peredur Webb-Davies, María del Carmen Parafita Couto, Draško Kaščelan

trices. Note that the authors of the study distinguish code-mixing from code-switching on the basis of pragmatic function. This chapter is noteworthy as the first corpus-based study into code-mixing in Heritage Piedmontese.

The authors draw their data (a sample of 22 speakers) from a larger corpus of spontaneous speech collected as part of an ongoing project focusing on the linguistic variety Heritage Piedmontese (HP). The corpus data includes semi-structured interviews between a native HP researcher and one or more HP speakers, with the group conversations involving more spontaneous participant-to-participant interaction. The authors highlight that the active use of HP among the community primarily found during specifically Piedmontese-related cultural activities, and the recorded conversations took place in an explicit monolingual mode; that is, participants were requested to use HP, though this did not preclude code-mixing of the type which is described by the authors.

They identify several distinct types of switching of Spanish noun material in their data, with insertion being much more common than alternation, although the authors note that this preference is not found to such an extent in the other data they collected (i.e., when not in an interview setting). The most frequently-observed insertional type in their data is what they call “minimal insertions”, i.e., the insertion of Spanish nouns either as bare forms (which comprises the majority of the tokens in their analysis, 64%) or with morphological Piedmontese morphology which serves to integrate switched words with the Piedmontese grammar used in the rest of the sentence. A significant difference was found for grammatical number, where plural nouns were most likely to be inserted with Piedmontese morphological marking than singular nouns, and a further significant difference was observed that contrasts inserted feminine Spanish nouns, which are more likely to be integrated to HP morphology, than masculine nouns, which are more likely to be bare forms.

The authors explain the patterns found as being emblematic of a preference by these participants to “display their knowledge” of Piedmontese and thereby avoid code-mixing where possible, a practice observed in other studies of similar social contexts (e.g., [Carter et al. \(2011\)](#) for Welsh-Spanish bilinguals in Argentina). Regarding the patterns shown with noun number and gender, the authors interpret this to relate to the grammatical congruence between the two languages (both Romance), i.e., at points of congruence between HP and Spanish, the same grammatical structure is maintained. A small number of instances of insertions are found which were not predicted by the authors, and which they suggest may indicate innovative forms.

The interaction between languages explored in this study can be classified as between a nationally prestigious language (Argentinian Spanish, the primary

1 *Code-switching in context: Variation, processing, and development*

language of that country) and a heritage language (Heritage Piedmontese, originating in Piedmont, Italy). Piedmontese is endangered in Piedmont and has restricted usage in Argentina as the language of heritage communities, though it has experienced a degree of language revival. Literature on situations involving speakers who are bilingual in a majority and a minority language has demonstrated the common influence of the former on the latter, such as in the form of language shift – characterized as the diachronic preference of speakers in a speech community for one language over another less prestigious one (e.g., [Appel & Muysken 2006](#)) – or linguistic interference (e.g., [Thomason & Kaufman 1988](#)). Case studies of the influence of code-switching in particular on grammatical changes within the minority language in such cases include the work of Deuchar and colleagues on Welsh-English bilinguals (e.g., [Deuchar & Davies 2009](#)). The study presented in this chapter provides a valuable addition to this field.

Additionally, the use of recorded spontaneous speech to analyse language use patterns is, of course, well-established in sociolinguistics. Numerous past projects have collated corpora of bilinguals' speech in a similar fashion, including, notably, the corpora of Welsh-English, Spanish-English and Welsh-Spanish corpora (the latter likewise from Argentina) by researchers based in Bangor University, Wales, under the auspices of Margaret Deuchar (q.v. [Deuchar et al. 2018](#)). That data is publicly accessible ([BangorTalk n.d.](#)), and indeed we view Margaret Deuchar as a pioneer in making bilingual linguistic data widely available for use. It is hoped that much more bilingual corpora are made openly accessible for researchers to verify and explore, especially taking into account considerable hard work involved in collecting, coding and transcribing corpus data in the first place (see, e.g., the Talkbank repository of corpus data under the stewardship of Brian MacWhinney ([MacWhinney 2000](#))); indeed, it is commendable that Gorla and Bellamy provided a link to their data via the Open Science Framework website.

Another corpus study, by Vihman and colleagues, is presented in chapter 3, with related data being made available on the Open Science Framework. The authors investigate code-switching in instant messaging (IM) by Estonian-English bilinguals aged under 18, focusing on the extent and nature of the morphological and orthographic integration of English elements into an Estonian frame. The teenagers under study live offline in a relatively monolingual Estonian sociolinguistic context, but engage with online chat in a much more multilingual manner ([Vihman et al. 2022](#)), using English particularly for stylistic or pragmatic function within their linguistic repertoire. The authors look at the test case of the language used in instant messaging sourced from a corpus that consists of IM from over 100 Estonian teenage participants; the study presented here includes

Peredur Webb-Davies, María del Carmen Parafita Couto, Draško Kaščelan

data from 99 participants. A clear contribution of this chapter is therefore its exploration of code-switching in computer-mediated communication, a mode that is text-based, and therefore governed to some extent by the orthographic conventions of the language in question, but is also frequently influenced by features of the spoken informal language, including not only phonological features but also code-switching, as the authors point out.

In general, Vihman et al. find a low proportion of English words in the dataset (7.5% of tokens; in contrast, speech data analyzed by Vihman et al. (2022) had 3% English tokens), and it is observed that these participants generally use an Estonian morphosyntactic frame for their IMs. The authors present a detailed quantitative analysis of orthographic and morphological treatment of English insertions, including examining participant variables and different parts of speech. Orthographic integration of English insertions into the Estonian frame was low: only 17% of tokens were either orthographically integrated or had a mixture of elements of English and Estonian orthography; the remainder were coded as English switches which retained English spelling. The results show some variation between participants based on factors including age, with younger participants integrating less than older ones. With regard to morphological integration, the authors note that most switched material in their data is not able to be integrated anyway, such as interjections and adverbs, but of the switches where morphological integration was available, the majority (79%) are indeed integrated, in contrast to the orthographic data, with less inter-participant variation. There are very few tokens in the dataset which exhibit both orthographic and morphological integration. The authors provide a detailed discussion of the interplay between orthographic and morphological treatment of English switching in the IM language of these teens, making a persuasive example of code-switching used for stylistic as well as communicative purposes.

The extent to which code-switched material from a language is integrated into the grammar of a recipient language has long been an interest of the field. Although some theoretical approaches argue that, e.g., the affixation of bound morphemes from one language onto a stem from another violates some constraint (cf. Poplack 1980), and indeed some studies find bare forms (e.g., Schmitt 2000); it has been demonstrated elsewhere that this occurs across different languages (e.g., Deuchar et al. 2018, Myers-Scotton 2002); see also Gorla and Bellamy in chapter 2 of this volume. Indeed, the Matrix Language Frame model (Myers-Scotton 2002, Myers-Scotton & Jake 2009) posits that it is to be expected that an inserted lexical item is morphologically integrated into the morphosyntactic frame into which it is embedded, as found in, e.g., Deuchar (2020) or Deuchar et al. (2018) – assuming, of course, that the languages in question have morphology that enables

1 *Code-switching in context: Variation, processing, and development*

this to be produced or observed. In the case of Estonian and English, Estonian is far more morphologically rich than English, having, for example, a complex system of inflectional case marking, and as such this provides an ideal context for identifying the extent of the integration of code-switched material.

The findings in Vihman et al.'s study here show that, while IM language is broadly comparable to speech norms (at least in their case), there is added interest in examining it as it presents other factors, such as spelling, to consider, but also because in text discourse, as the authors demonstrate here, it is possible to evaluate more closely the apparent intent of users in how they are deploying code-switched material, e.g., via the extent to which they integrate the orthography and what this implies. We welcome the expansion of the study of code-switching into non-spoken contexts and look forward to more contributions like this.

In chapter 4, Ezeizabarrena, Munarriz-Ibarrola and deCastro-Arrazola describe their study of Basque-Spanish code-switching, focusing on the interaction of Basque and Spanish within determiner phrases (DP). So-called mixed DPs have been, as the authors note, the focus of several previous studies, including such language pairs as English-Spanish (Dussias 2003, Valdés Kroff 2016) and Welsh-English (Parafita Couto et al. 2013, Parafita Couto & Gullberg 2019); see Bellamy & Parafita Couto 2022 for an overview. In the case of languages with different gender rules – and, e.g., different ways of marking gender within DPs – this is a particularly interesting focus for analysing code-switching, since speakers producing such phrases must select some combination of morphemes from the two interacting languages, which may or may not adhere to the grammar of one or both those languages. In the case of Basque and Spanish, Spanish nouns have gender morphology whereas Basque does not; furthermore, there are other differences with the behaviour of elements such as determiners between the two languages. Previous studies have identified varying patterns of preference by judges on how Basque nouns interact with Spanish determiners (Badiola & Sande 2018, Iriondo Etxeberria 2017, Parafita Couto et al. 2015), which Munarriz-Ibarrola et al. (2021) demonstrate is linked to a bilingual's first language. Ezeizabarrena and colleagues elicit grammaticality judgments from two groups of Basque-Spanish bilinguals, L1 Spanish (also comprising simultaneous bilinguals and early sequential bilinguals) and L1 Basque (all early sequential bilinguals). Their general aim was to identify which patterns of integration of Basque nouns in mixed DPs are preferred by the participants.

The method implemented in this study required participants to listen to stimuli sentences containing mixed Basque-Spanish DPs comprising a Spanish determiner preceding a Basque noun, and to rate their acceptability on a scale from

Peredur Webb-Davies, María del Carmen Parafita Couto, Draško Kaščelan

1 (bad) to 5 (good). The results show a mean acceptability of 3.54 for the critical sentences, which the authors interpret as general acceptance or familiarity with mixed DPs by these speakers, although participants did not rate such sentences as highly as unilingual Spanish or Basque sentences. The judgment patterns of both participant groups were broadly similar, although the L1 Spanish bilinguals' responses were qualitatively more polarised than those of the L1 Basque bilinguals. The authors conduct statistical tests that identify some patterns that contrast the two groups' responses, though no clear overall preference was shown to grammatical gender in mixed DPs, in contrast to previous studies. The authors conclude that their findings are indicative of speakers using word-specific strategies when it comes to gender marking in mixed DPs, or that certain kinds of bilingual (e.g., relating to language dominance) may prefer certain strategies.

Valdés Kroff et al.'s chapter 5 in this volume, following up on one of their previous studies (Valdés Kroff et al. 2017), report on their investigation into an established finding in the literature on Spanish-English code-switching, whereby switch points that occur before an auxiliary such as *haber* 'have' are found much more commonly than switch points after such an auxiliary. Recruiting L1 English-L2 Spanish participants living in Florida, the authors aimed to test participants' reactions to code-switched sentences of varying degrees of grammaticality (i.e., including both sentence types that are typically found in the speech of such bilinguals, as well as those that are seldom or never found). Spanish-English bilinguals in Florida have been the focus of numerous studies (e.g., Parafita Couto et al. 2010, Parafita Couto et al. 2013); speakers in this population have been argued to index code-switching with identity and to include both inter-sentential and intra-sentential switching in their repertoires (e.g., Carter et al. 2011). As a population that, varying by city or region, is intensely exposed to both Spanish and English – and Spanish-English code-switching – in their daily lives, such bilinguals are well-suited candidates for investigating the practice of and responses to code-switching from a psycholinguistic perspective.

In the present work by Valdés Kroff et al., participants undertook an eye-tracking experiment requiring them to read test sentences which included code-switching between Spanish and English, but which varied in terms of the switch points within the sentence: *haber* 'have' + verb participle constructions with a switch either before or after *haber*, and *estar* 'be' + verb participle constructions with a switch either before or after *estar*; sentences were also divided between perfect and present tense constructions. The participants were further contrasted based on what task they were asked to complete after reading each sentence: either to judge whether the sentence was grammatical or not, or to indicate whether or not they comprehended the sentence.

1 Code-switching in context: Variation, processing, and development

A key finding of this study is a difference between the group who did the comprehension task and those who did the grammaticality judgment task. Based on the early reading measure of gaze duration, the group participating in the comprehension task was found to read switches at the participle more slowly than at the auxiliary (with both auxiliary types - *haber* and *estar*). However, based on the total duration, this group showed slower reading times at the participle than at the auxiliary when *haber* was used, while there were no differences in relation to auxiliary *estar*. On the other hand, the participant group who did the grammaticality judgment task did not demonstrate the same extent of variability in reading speed: for this group, overall trials on which switches occurred at the participle were read more slowly than at the auxiliary. This distinction between the two groups may be interpreted as evidence that being asked whether a sentence is grammatical is a metalinguistic task which represents natural language processing less well than when asked whether a sentence is meaningful, based on the evidence from the comprehension task group, whose responses are similar to the patterns already described in the literature.

The authors argue that the findings of their study point to the online processing of code-switching matching speech norms of the community; e.g., since a switching distinction between *haber* and *estar* is typical in speech among Spanish-English bilinguals in Florida (for example), that distinction is also reflected in the psycholinguistic processing by such bilinguals. The authors rightly point out that code-switching norms vary from community to community (q.v. Blokzijl, Deuchar, and Blokzijl et al. 2017, Carter et al. 2011, Deuchar 2020), which is indeed motivation for further replication of their method to other contrastive speech communities, in order to see whether the findings would be repeated in populations where, for example, the languages in question or the code-switching habits of speakers are different.

This is a valuable study that further helps to bridge research in language from a social perspective and language from a psycholinguistic perspective: cf. earlier forays into this by Vaughan-Evans et al. (2020), the authors of which were a mix of psychologists and linguists; that study also found, in broad terms, a parallel between electrophysiological evidence and the code-switching patterns found in the speech of Welsh-English bilinguals (e.g., Deuchar et al. 2018).

1.2 Development and neurodevelopmental conditions

Several chapters in this volume explore code-switching (predominantly) in bilingual children to examine the critical role of the surrounding linguistic environment. These chapters illuminate the ways in which community language, care-

Peredur Webb-Davies, María del Carmen Parafita Couto, Draško Kaščelan

giver practices, and prevailing attitudes toward bilingualism converge to shape the code-switching patterns of young children, both neurotypical and neurodivergent.

Earlier approaches to code-switching have often focused on linguistic constraints, seeking to identify the grammatical principles that govern language mixing. While such analyses are valuable, they often overlook the dynamic interplay between the individual and their environment (for an overview see [Deuchar 2020](#), [Parafita Couto et al. 2023](#), [Phillips & Deuchar 2022](#)). The studies in this volume place code-switching within a broader ecological framework, recognizing that language is not simply a cognitive tool but a social practice embedded in specific communities and relationships. The ecological perspective adopted here highlights the interplay between individual linguistic development and broader socio-cultural contexts.

In chapter 6, Vihman and Vihman review and analyse 47 studies on the spontaneous language use by bilingual children between the ages of 2-4 years. By looking at emergent grammars, they found that almost every aspect of early morphosyntax (apart from root infinitives) is open to the influence of bilingual interaction. In terms of type of interaction, they find evidence of acceleration, delay, and transfer between structures. The review remains inconclusive about the effect that dominance plays in interaction between structures, likely due to various ways in which dominance has been operationalised across studies. Vihman and Vihman point to the importance of individual differences in dynamic experiences of bilinguals, which might affect whether dominance comes up as an important factor. This review adds to the long standing discussions in the field of bilingualism about the lack of consensus on how specific phenomena, such as dominance, are defined (e.g., [Treffers-Daller 2019](#)), which ultimately has an impact on comparisons between existing research and generalisations that can be drawn about bilingual development. Vihman and Vihman conclude that emerging bilingual grammars are moving targets - consequently, it might be hard to estimate the required degree of overlap between languages to trigger interaction. As the previous overviews of bilingual grammars focused on experimental studies and included older children (e.g., [Van Dijk et al. 2022](#)), Vihman and Vihman's overview supplements their work by looking at spontaneous speech of emerging bilinguals. As they point out, small samples across identified 47 studies illustrate a need for large scale research within this age group. In light of that, their review is a vital state-of-the-art summary which maps out areas that require further attention, such as looking at patterns of use in the wider community and their effect on language use in young bilinguals.

1 *Code-switching in context: Variation, processing, and development*

Work by Quay (chapter 7) illustrates the value of an in-depth case study approach in understanding the dynamic nature of bilingual development in children, as well as the impact of caregivers, such as grandparents. The case study follows the linguistic development of Ryan, a Cantonese-English bilingual, between the ages of 2;9 and 4;11 in his interactions with his grandmother. Sixteen transcripts (totalling 460 minutes of video recordings) in which Ryan and his grandmother were alone were analysed to explore changes in his language use patterns. The study aimed to explore whether introducing more English (the societal language) through code-switching practices would facilitate Ryan's frequency of use and skill development in English (his 'weaker' language), and his overall development as a Cantonese-English bilingual in the Canadian context. Over the course of 2 years, data indicated that Ryan is generally accommodating his grandmother's input in terms of the frequency patterns of using English-only, Cantonese-only, or code-switched utterances. In terms of intra-sentential code-switching (i.e., switching within a sentence), Ryan and his grandmother had a similar proportion of one-word insertions (62% and 69% respectively), two-word mixed utterances (8% each), and multi-word mixed utterances (30% and 23% respectively). Further qualitative analyses identified that Ryan's English utterances were morphosyntactically less complex than his Cantonese utterances, suggesting that despite accommodating his grandmother's language choice, he remained a Cantonese-dominant bilingual. Unlike his grandmother's input, Ryan's speech showed evidence of congruent lexicalisation and convergence, which he used in more complex and in compound sentences, as well as to form complete utterances in his 'weaker' language. Quay argues that this data stresses the importance of analysing single-language utterances by emerging bilinguals, especially in their 'weaker' language, as they can offer insight into language mixing phenomena through convergence. This longitudinal observation shows how grandmother's code-switching practices as well as the child's pragmatic sensitivity to language choice facilitated the development of initially 'weaker' language (i.e., the societal language) of an emerging bilingual. Importantly, this data highlights that unrestricted use of both languages facilitates linguistic development of bilinguals as it allows drawing upon their complete linguistic repertoire.

The scoping review by Kaščelan and Parafita Couto (chapter 8) outlines existing findings on code-switching by individuals with neurodevelopmental conditions. By looking at this under-explored area, they identify lack of representation in various aspects. Specifically, in their analysis of 31 manuscripts, the focus was primarily on code-switching in autism and language disorder, while code-switching in other neurodevelopmental conditions (e.g., ADHD, stuttering, etc.) has been rarely investigated. In terms of demographic-related data, most stud-

Peredur Webb-Davies, María del Carmen Parafita Couto, Draško Kaščelan

ies focused on children of age 8 or under, and with participants predominantly from North America and Europe. Code-switching in geographic locations where multilingualism and code-switching tend to be mainstream norms of communication are rarely explored. The most striking finding of this review points to the approaches in which code-switching has been explored and their ecological validity. In particular, only 6 studies relied on spontaneous speech data, while 4 other studies included a combination of spontaneous speech and (semi-)experimental approaches. Finally, Kaščelan and Parafita Couto identify lack of research in attitudes and advice that bilingual families with neurodivergent family members receive in terms of their code-switching practices. On rare occasions when this data is available, the advice and attitudes are predominantly negative, thus reinforcing stigmatisation of bilingual speech.

Their chapter offers recommendations for future research, highlighting the importance of slow science for better understanding of code-switching by neurodivergent individuals. Importantly, in line with recommendations by Deuchar (2020) among others, availability of spontaneous speech data is crucial in our shared understanding of this phenomenon, especially in cases such as neurodevelopmental conditions where the number of participants across existing studies has been particularly small. Making data available would allow for aggregated dataset creation, it would enable comparisons between different bilingual communities, and contribute to the validity of results. Advice and implications for educators and speech and language therapists are also offered based on the current findings in this field.

The contributions in this section collectively illustrate the intricate relationship between the characteristics of the languages, individual characteristics, caregiver strategies, and broader community practices in shaping bilinguals' code-switching patterns.

2 Concluding remarks

The structure of this volume mirrors the intellectual trajectory and influence of Margaret Deuchar in the study of code-switching. Beginning with investigations into grammatical variation and constraints on code-switching, the volume reflects Deuchar's pioneering contributions to the linguistic patterns governing bilingual speech (e.g., Deuchar 2006, Deuchar et al. 2007). Her work on Welsh-English, Spanish-English and Spanish-Welsh bilingualism and corpus-based methodologies laid the foundation for understanding how different languages interact at a structural level (BangorTalk n.d., Deuchar 2020).

1 Code-switching in context: Variation, processing, and development

Furthermore, this volume extends Deuchar's legacy by emphasizing how code-switching operates in dynamic linguistic and social contexts. Deuchar's more recent work has increasingly incorporated interdisciplinary perspectives, bridging psycholinguistics, sociolinguistics, and corpus linguistics (e.g., [Vaughan-Evans et al. 2020](#)). This is evident in the studies presented here, which explore code-switching as a product of both cognitive processes and social interaction, in line with the broader scope she championed.

Finally, the inclusion of research on neurodevelopmental conditions highlights an emerging frontier in bilingualism studies, one that aligns with Deuchar's advocacy for empirical, data-driven research and her commitment to making bilingual linguistic data widely accessible. The emphasis on spontaneous speech corpora and ecological validity throughout the volume echoes her pioneering role in corpus creation and open data initiatives (e.g., [Deuchar et al. 2014, 2018](#)).

In sum, this volume is not just a collection of studies on code-switching; it is a reflection of the evolution of the field itself, shaped in no small part by Deuchar's contributions. Her influence permeates the chapters, reinforcing the idea that code-switching is a window into the complexities of bilingualism/multilingualism, one that continues to reveal new insights, just as her work has done over the decades.

References

- Appel, Rene & Pieter C. Muysken. 2006. *Language contact and bilingualism*. Amsterdam: Amsterdam University Press. <https://www.aup.nl/en/book/9789048504138/language-contact-and-bilingualism>.
- Badiola, Lucia & Ariane Sande. 2018. Gender assignment in Basque/Spanish mixed Determiner Phrases: A study of simultaneous bilinguals. In Luis López (ed.), *Code-switching – Experimental answers to theoretical questions: In honor of Kay González-Vilbazo*, 15–38. Amsterdam: John Benjamins.
- BangorTalk. N.d. *BangorTalk bilingual conversational corpora*. <https://bangortalk.org.uk/>.
- Bellamy, Kate & Maria del Carmen Parafita Couto. 2022. Gender assignment in mixed noun phrases: State of the art. In Dalila Ayoun (ed.), *The acquisition of gender: Crosslinguistic perspectives* (Studies in Bilingualism 63), 13–48. Amsterdam: John Benjamins. DOI: [10.1075/sibil.63.02bel](https://doi.org/10.1075/sibil.63.02bel).
- Blokzijl, Jeffrey, Margaret Deuchar & Maria del Carmen Parafita Couto. 2017. Determiner asymmetry in mixed nominal constructions: The role of grammatical factors in data from Miami and Nicaragua. *Languages* 2(4). 20. DOI: [10.3390/languages2040020](https://doi.org/10.3390/languages2040020).

Peredur Webb-Davies, María del Carmen Parafita Couto, Draško Kaščelan

- Carter, Diana, Peredur Davies, Margaret Deuchar & María del Carmen Parafita Couto. 2011. A systematic comparison of factors affecting the choice of matrix language in three bilingual communities. *Journal of Language Contact* 4(2). 153–183. DOI: [10.1163/187740911X592808](https://doi.org/10.1163/187740911X592808).
- Deuchar, Margaret. 2006. Welsh-English code-switching and the Matrix Language Frame model. *Lingua* 116(11). Celtic Linguistics. 4th Celtic Linguistics Conference, 1986–2011. DOI: [10.1016/j.lingua.2004.10.001](https://doi.org/10.1016/j.lingua.2004.10.001).
- Deuchar, Margaret. 2020. Code-switching in linguistics: A position paper. *Languages* 5(2). 22. DOI: [10.3390/languages5020022](https://doi.org/10.3390/languages5020022).
- Deuchar, Margaret & Peredur Davies. 2009. Code switching and the future of the Welsh language. *International Journal of the Sociology of Language* 2009(195). 15–38. DOI: [10.1515/IJSL.2009.004](https://doi.org/10.1515/IJSL.2009.004).
- Deuchar, Margaret, Peredur Davies, Jon Russell Herring, Maria del Carmen Parafita Couto & Diana Carter. 2014. Building bilingual corpora. In Enlli Môn Thomas & Ineke Mennen (eds.), *Advances in the study of bilingualism*, 93–110. Bristol: Multilingual Matters. DOI: [10.21832/9781783091713-008](https://doi.org/10.21832/9781783091713-008).
- Deuchar, Margaret, Pieter Muysken & Sung-Lan Wang. 2007. Structured variation in codeswitching: Towards an empirically based typology of bilingual speech patterns. *International Journal of Bilingual Education and Bilingualism* 10(3). 298–340. DOI: [10.2167/beb445.0](https://doi.org/10.2167/beb445.0).
- Deuchar, Margaret, Peredur Webb-Davies & Kevin Donnelly. 2018. *Building and using the Siarad Corpus: Bilingual conversations in Welsh and English* (Studies in Corpus Linguistics 81). Amsterdam: John Benjamins. DOI: [10.1075/scl.81](https://doi.org/10.1075/scl.81).
- Dussias, Paola E. 2003. Spanish-English code mixing at the Auxiliary Phrase: Evidence from eye-movement data. *Revista Internacional de Lingüística Iberoamericana* 1(2). 7–34. <https://www.jstor.org/stable/41678169>.
- Iriondo Etxeberria, Alejandra. 2017. *Kode-alternantzia elebidun gazteetan: DS nahiak [(Code switching in young bilinguals: Mixed DP)]*. Bilbao: University of the Basque Country. (Doctoral dissertation). DOI: [10.1387/gogoa.18646](https://doi.org/10.1387/gogoa.18646).
- MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk: Transcription format and programs*. 3rd edn. Mahwah: Lawrence Erlbaum Associates Publishers.
- Munarriz-Ibarrola, Amaia, Maria-José Ezeizabarrena, Varun D. C. Arrazola & Maria del Carmen Parafita Couto. 2021. Gender assignment strategies and L1 effects in the elicited production of mixed Spanish-Basque DPs. *Linguistic Approaches to Bilingualism* 12(6). 778–815. DOI: [10.1075/lab.20016.mun](https://doi.org/10.1075/lab.20016.mun).
- Muysken, Pieter. 2000. *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.

1 *Code-switching in context: Variation, processing, and development*

- Myers-Scotton, Carol (ed.). 2002. *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press. DOI: [10.1093/acprof:oso/9780198299530.002.0003](https://doi.org/10.1093/acprof:oso/9780198299530.002.0003).
- Myers-Scotton, Carol & Janice L. Jake. 2009. A universal model of code-switching and bilingual language processing and production. In Almeida Jacqueline Toribio & Barbara E. Bullock (eds.), *The Cambridge handbook of Linguistic Code-switching* (Cambridge Handbooks in Language and Linguistics), 336–357. Cambridge: Cambridge University Press. DOI: [10.1017/CBO9780511576331.020](https://doi.org/10.1017/CBO9780511576331.020).
- Parafita Couto, Maria del Carmen, Amaia Munarriz, Irantzu Epelde, Margaret Deuchar & Beñat Oyharçabal. 2015. Gender conflict resolution in Spanish–Basque mixed DPs. *Bilingualism: Language and Cognition* 18(2). 304–323. DOI: [10.1017/S136672891400011X](https://doi.org/10.1017/S136672891400011X).
- Parafita Couto, Maria del Carmen, Kate Bellamy & Felix K. Ameka. 2023. Theoretical linguistic approaches to multilingual code-switching. In Jennifer Cabrelli, Adel Chaouch-Orozco, Jorge González Alonso, Sergio Miguel Pereira Soares, Eloi Puig-Mayenco & Rothman Jason (eds.), *The Cambridge handbook of third language acquisition* (Cambridge Handbooks in Language and Linguistics), 403–436. Cambridge: Cambridge University Press. DOI: [10.1017/9781108957823.017](https://doi.org/10.1017/9781108957823.017).
- Parafita Couto, Maria Del Carmen, Diana Carter, Peredur Davies & Margaret Deuchar. 2010. Factors determining code-switching patterns in Spanish-English and Welsh-English communities. In Jorge Luis Bueno Alonso (ed.), *Analizar datos > Describir variación. Analyzing data > Describing variation. XXVIII Congreso Internacional de la Asociación Española de Lingüística Aplicada (AESLA)*, 94. Vigo: Universidade de Vigo.
- Parafita Couto, Maria del Carmen, Peredur Davies, Diana Carter & Margaret Deuchar. 2013. Factors influencing code-switching. In Enlli Thomas & Ineke Mennen (eds.), *Advances in the study of bilingualism*, 111–140. Bristol: Multilingual Matters.
- Parafita Couto, Maria del Carmen & Marianne Gullberg. 2019. Code-switching within the noun phrase: Evidence from three corpora. *International Journal of Bilingualism* 23(2). 695–714. DOI: [10.1177/1367006917729543](https://doi.org/10.1177/1367006917729543).
- Phillips, Shannon & Margaret Deuchar. 2022. The role of the input in the acquisition of code-switching. In Unn Røyneland & Robert Blackwood (eds.), *Multilingualism across the lifespan* (Routledge Critical Studies in Multilingualism), 56–79. New York: Routledge. DOI: [10.4324/9781003125815-5](https://doi.org/10.4324/9781003125815-5).
- Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics* 18(7/8). 581–618. DOI: [10.1515/ling.1980.18.7-8.581](https://doi.org/10.1515/ling.1980.18.7-8.581).

Peredur Webb-Davies, María del Carmen Parafita Couto, Draško Kaščelan

- Schmitt, Elena. 2000. Overt and covert codeswitching in immigrant children from Russia. *International Journal of Bilingualism* 4(1). 9–28. DOI: [10.1177/13670069000040010201](https://doi.org/10.1177/13670069000040010201).
- Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press. <https://www.ucpress.edu/books/language-contact-creolization-and-genetic-linguistics/paper>.
- Treffers-Daller, Jeanine. 2019. What defines language dominance in bilinguals? *Annual Review of Linguistics* 5. 375–393. DOI: [10.1146/annurev-linguistics-011817-045554](https://doi.org/10.1146/annurev-linguistics-011817-045554).
- Valdés Kroff, Jorge R., Paola E. Dussias, Chip Gerfen, Lauren Perrotti & M. Teresa Bajo. 2017. Experience with code-switching modulates the use of grammatical gender during sentence processing. *Linguistic Approaches to Bilingualism* 7(2). 163–198. DOI: [10.1075/lab.15010.val](https://doi.org/10.1075/lab.15010.val).
- Valdés Kroff, Jorge R. 2016. Mixed NPs in Spanish-English bilingual speech: Using a corpus-based approach to inform models of sentence processing. In Rosa E. Guzzardo Tamargo, Catherine M. Mazak & Maria del Carmen Parafita Couto (eds.), *Spanish-English codeswitching in the Caribbean and the US* (Issues in Hispanic and Lusophone Linguistics 11), 281–300. Amsterdam: John Benjamins. DOI: [10.1075/ihll.11.12val](https://doi.org/10.1075/ihll.11.12val).
- Van Dijk, Chantal, Elise Van Wonderen, Elly Koutamanis, Gerrit Jan Kootstra, Ton Dijkstra & Sharon Unsworth. 2022. Cross-linguistic influence in simultaneous and early sequential bilingual children: A meta-analysis. *Journal of Child Language* 49(5). 897–929. DOI: [10.1017/S0305000921000337](https://doi.org/10.1017/S0305000921000337).
- Vaughan-Evans, Awel, Maria del Carmen Parafita Couto, Bastien Boutonnet, Noriko Hoshino, Peredur Webb-Davies, Margaret Deuchar & Guillaume Thierry. 2020. Switchmate! An electrophysiological attempt to adjudicate between competing accounts of adjective-noun code-switching. *Frontiers in Psychology* 11. DOI: [10.3389/fpsyg.2020.549762](https://doi.org/10.3389/fpsyg.2020.549762).
- Vihman, Virve-Anneli, Kristiina Praakli, Maarja-Liisa Pilvik & Mari-Liis Korus. 2022. Kas noored on inglise keelele ‘obsessed’? Millest räägivad korpusandmed? [Are young people “obsessed” with English? What do corpus data tell us?] *Philologia Estonica Tallinnensis* 7. 292–321. DOI: [10.22601/PET.2022.07.11](https://doi.org/10.22601/PET.2022.07.11).

Chapter 2

Code-mixing and language change: Evidence from Spanish nouns in heritage Piedmontese

Eugenio Gorla & Kate Bellamy

In this paper we provide a qualitative and quantitative analysis of Argentinian Spanish noun insertion in Heritage Piedmontese (HP henceforth), an Italo-Romance minority language that is also spoken in Argentina. Despite language shift towards the national language, Spanish, HP continues to be used in a number of culturally relevant situations. In these situations, where code-mixing is commonplace, HP is usually the matrix language of bilingual interaction. We show that mixed noun phrases (NPs) display considerable variation in the realisation of gender and number marking, despite a certain amount of morphological and phonological congruence between the nominal systems of the two languages. Our analysis is based on a sample of five hours of speech, extracted from a larger corpus of interviews and conversations in HP collected between 2019 and 2022 in the framework of the PILAR - Piedmontese Language in Argentina project. Building on previous taxonomies, we identify three main types of code-mixing in the HP-Spanish data: (i) alternations, where a Spanish NP is part of a code-switch spanning over several constituents; (ii) maximal insertions, involving a Spanish embedded-language island; and (iii) minimal insertions, where a Spanish noun is inserted into a Piedmontese morphological structure, either as a bare stem or adopting Piedmontese morphology. We identified 318 mixed NPs, of which the bare stem type is the most common. On the basis of a comparison of gender and number marking in Spanish and Piedmontese, we demonstrate that the observed types of code-mixing are mostly predicted by typological similarities and differences between the two languages involved. Also new patterns of number marking have emerged from the analysis, which suggest a new hypothesis on the relationship between code-mixing and language change.

Eugenio Gorla & Kate Bellamy

1 Introduction

In this paper we analyse code-mixing¹ in the Piedmontese-speaking community in Argentina. Piedmontese is an Italo-Romance minority language, spoken in the northwestern Italian region of Piedmont (Italian: Piemonte). It does not have official status in Italian law, and has undergone normative standardisation only to a certain degree (see [Regis 2013](#), [Regis & Rivoira 2023](#), for a more detailed discussion). Piedmontese was regularly used, especially in informal situations, until the first half of the twentieth century, but is currently undergoing language shift towards Standard Italian along the well-described path from spoken diglossia to diaglossia ([Auer 2005](#)), or ‘dilalia’ ([Berruto 2012](#)). This shift describes a pattern in the evolution of linguistic repertoires that is common to most European countries: national standards, after starting out as H[igh] varieties in a diglossic situation, and thus almost exclusively confined to written uses, increasingly extend to the L[ow] domains, typical of spoken vernaculars, eventually becoming the main language within the household. A case in point is Standard Italian, which evolved from the status of an almost exclusively written to a spoken national language throughout the twentieth century, due to social and historical reasons, such as the introduction of compulsory primary education, the diffusion of Italian-speaking television, and internal migrations ([De Mauro 1963](#)). This resulted in a new configuration of the linguistic repertoire, where Piedmontese and the other Romance vernaculars spoken in Italy compete with the national language even in the most familiar and informal domains, thereby reducing the vernaculars to optional codes.²

As an effect of Italian emigration, and especially during the period between the end of the nineteenth century and the outbreak of World War I, Piedmontese also came to be spoken outside of Italy, and particularly in Argentina, where the socio-political, as well as territorial conditions were favourable to its retention among the early migrants, as well as by their descendants (see [Giolitto 2000, 2010, 2016](#), [Gorla 2015, 2021](#), [Doglioli 2022](#)). We will refer to this variety throughout this paper as Heritage Piedmontese (HP), because both its linguistic features and the social conditions in which it is spoken reflect the typical features of heritage languages. We thus follow Jason Rothman’s definition of a heritage language, whereby:

¹The term ‘code-mixing’ is used here following [Auer \(1999\)](#) as a type of bilingual speech phenomenon where the juxtaposition between two (or more) codes in the same discourse does not bear any pragmatic meaning. See Section 2.

²Differences between Auer’s notion of ‘diaglossia’ and Berruto’s ‘dilalia’ concern the status of the L variety, which is fundamentally a regiolect of H in Auer’s account, but a separate linguistic system functionally subordinated to H (that is, a minority language) in Berruto’s account.

2 *Code-mixing and language change*

A language qualifies as a heritage language if it is a language spoken at home or otherwise readily available to young children, and crucially this language is not a dominant language of the larger (national) society [...] the heritage language is acquired on the basis of an interaction with naturalistic input and whatever in-born linguistic mechanisms are at play in any instance of child language acquisition. Differently [from monolingual acquisition], there is the possibility that quantitative and qualitative differences in heritage language input, influence of the societal majority language and differences in literacy and formal education can result in what on the surface seems to be arrested development of the heritage language or attrition in adult bilingual knowledge (Rothman 2009: 156)

Based on recent research on heritage languages (see, e.g., Benmamoun et al. 2013, Nagy 2017, Polinsky 2018, Aalberse et al. 2019, Gorla & Di Salvo 2023) we may thus expect Heritage Piedmontese to be shaped on the one hand by acquisitional factors, and on the other hand by contact with local Spanish as the dominant language in society.³ Indeed, given the long-term contact with Spanish and its resulting bilingualism, it is perhaps unsurprising that many individuals, especially those in the provinces of Córdoba and Santa Fé, where Piedmontese migration has been particularly intensive, use both of their languages in the same conversation, sentence or phrase. We will refer to this practice as ‘code-mixing’; since the same term has been used by different authors and in different theoretical frameworks, we will follow, more specifically, the distinction made by Auer (1999, 2014), where code-mixing (as opposed to code-switching) is a type of bilingual speech where two languages are juxtaposed in discourse in a patterned way, and which is also often globally meaningful as a marker of mixed social identities (see also Footnote 1).

The rest of this chapter is structured as follows. In Section 2, we discuss our approach to code-mixing and introduce the categories for classifying nominal insertion that were used in this study. In Section 3 we provide some general context on HP in Argentina, and we introduce the data and methods that were used for this paper. In Section 4, based on a comparison of nominal inflection in HP and Spanish, we discuss our research questions and predictions. Results are presented in Section 5, and in Section 6 we draw some conclusions and directions for future research.

³A third dimension in this respect would be ‘internal’ contact between different varieties of Piedmontese, with the formation of a migrant koiné (see Kerswill 2006) as one of the possible outcomes. The possibility is discussed in Cerruti et al. (2025) and will not be further discussed here.

Eugenio Gorla & Kate Bellamy

2 On code-mixing

Code-mixing, or language mixing, constitutes the “juxtaposition of two languages in which the use of two languages is meaningful (to participants) not in a local but only in a more global sense, that is, when seen as a recurrent pattern” (Auer 1999: 310). Within this framework, code-mixing is opposed to code-switching as the latter reflects the strategic use of multiple languages in interaction and is typically associated with discourse-related or participant-related functions.

Code-mixing may be insertional in nature, where an element or elements from one language may be “inserted” into a main, or matrix, language (see also Myers-Scotton 1993, 2002, Muysken 2000); or it may be alternational, where (short) stretches of one language are juxtaposed with stretches of the other (or others), making it difficult to define a matrix language (see also Muysken 2000, who indeed refers to this as “alternation”).⁴ Typical examples of alternation and insertion are given in examples (1) and (2)–(3) respectively. All examples are from the first author’s corpus, unless otherwise stated; as a convention, throughout the paper we will use lower-case for Piedmontese and capitals for Spanish.

- (1) ENTONCES, a=smij-a che a=l’=han brusa=se ij
so 3SG=look-3SG that 3SG=CLIT=have:3PL burn=REFL the:M.PL
PAPEL EN LA ADUANA
papers in the customs
‘it looks like the documents were burnt at the customs’
- (2) j=er-o contadin da_là CHACAR-é da_sì
CLIT=be-3PL peasant there peasant-DERIV here
‘they were peasants there and peasants here’
- (3) perché MI pare parl-o nen LA HISTORI-A SUY-A en
because my parent speak-3PL NEG the:F.SG story(F)-SG POSS.3PL-F.SG in
piemonteis
Piedmontese
‘because my parents did (lit. do) not tell their story in Piedmontese’

⁴Muysken’s (2000) typology includes a third type of mixing, that is congruent lexicalisation. This implies the presence of a shared syntactic structure that complies with the abstract rules of both languages, and is filled with lexical items from either language giving rise to rather complex switches. Since this paper mostly deals with insertion, we did not consider this third type.

2 Code-mixing and language change

Based on Muysken's (2000, 2013) taxonomy and further elaborations such as Deuchar et al. (2007), example (1) represents a typical case of alternation, as the Spanish part of the sentence spans over several constituents and does not form a constituent itself. Rather it can be said that the sentence begins in Piedmontese and continues in Spanish. Examples (2) and (3) represent two different cases of insertion, which means that in both cases, Spanish nouns are inserted in an otherwise Piedmontese syntactic structure. In (2) the speaker is referring to the fact that Piedmontese migrants in Argentina had a similar occupation before and after migration, but it was given a different name. So, the lexical morpheme of the Argentinian Spanish *chacarero* 'peasant' is used in combination with the Piedmontese derivational morpheme *-é* which typically forms agent nouns (e.g. *verdur-a* 'groceries' > *verdur-é* 'greengrocer').⁵ According to Auer (2014), this constitutes a minimal insertion, as a single lexical morpheme from one language is embedded within the grammatical frame of another language. Conversely, in (3) *la historia suya* 'their story' (lit. 'the story their') constitutes a maximal insertion in Auer's terms, as a full NP in Spanish is inserted into an otherwise Piedmontese clause, and follows the rules of Piedmontese. Not only does the inserted NP contain solely lexical material from the embedded language (that is, Spanish), but it is also internally constructed following the syntactic rules of Spanish: homeland Piedmontese would have a different word order in this construction, namely a prenominal possessive determiner and no definite article (e.g. *sua stòria* 'their story'). Auer (2014) refers to these maximal insertions also as embedded-language islands, following Myers-Scotton (2002).

While the three code-mixing patterns described here appear to be easily distinguishable from each other, the interpretation of corpus data is not always so straightforward, and only in a few restricted cases is the distinction between the types categorical. Rather, most cases are best modelled as nodes on a continuum between insertion and alternation. Since the aim of this paper is to provide an analysis of nominal insertion in Heritage Piedmontese, we propose here a more fine-grained classification of insertional code-mixing that enables us to provide a more thorough analysis of this phenomenon. Thus, in order to classify the bilingual patterns identified in the corpus, we propose the categories that are summarised in Table 1.

Since we are relying on a corpus of interviews and data collected in situations of language revitalisation (see Section 3 for further discussion), HP tends to be firmly established as the base language in interaction and the matrix language in code-mixing. Based on this general consideration, we regarded as instances

⁵The examples are from the first author's personal knowledge.

Eugenio Gorla & Kate Bellamy

Table 1: Criteria adopted for the categorisation of bilingual patterns

Extension of the switch	Label	Type of morphological marking within the Noun Phrase	Example no.
> 1 constituent	Alternation	Spanish morphology	(1)
> 1 morpheme within the same constituent	Maximal insertion	Spanish morphology	(3), (4), (5)
only lexical morpheme	Minimal insertion	Bare stems	(6)
		Matrix language inflection	(2), (7)

of alternation proper all the cases like (1), where multiple elements in Spanish, which do not form a syntactic constituent, are juxtaposed in otherwise HP discourse. Following Auer (2014), we then identified a major distinction between maximal and minimal insertions. Maximal insertion in our view includes all the bilingual patterns where more than one morpheme is involved from the embedded language, among which we identify at least: complex NPs, involving several words, for example determiners, adjectives etc., as in (3); lists and other types of syntactically heavy constituents, as in (4); and single inflected nouns retaining Spanish overt morphology, as in (5).

- (4) prima fas-ÍAMO COSECHA e dòpo TAMBO LECHE
 before do-IMPF:1PL HARVEST and then DAIRY MILK
 ‘In the past we used to do harvest, and then dairy farm, milk (products)’
- (5) soma part-ì quandì l’=oma program-à ël
 be:AUX.PLL leave-PP when CLIT=have:AUX.1PL program-PP the:M.SG
 viaggi ch- a vint PESO-S
 travel that at twenty peso-PL
 ‘We left when we planned our travel, that (one dollar was) twenty PESOS’

Finally, minimal insertions are represented by cases where only the lexical morpheme is inserted into the other-language grammatical frame. In the case

2 Code-mixing and language change

of HP-Spanish code-mixing, this may consist of two different patterns. The first one is represented by the insertion of bare nouns, that is stems that have no overt morphological marking in the donor language (as is the case with Spanish singular nouns), or which are treated as invariables in bilingual discourse, thereby losing their morphological marking; consider example (6).

- (6) *ël pì cit dij VARÓN*
 the:M.SG more young of.the:M.PL male
 ‘The youngest of the males’

In (6), the Spanish word *varón* ‘male’ is inserted into an HP clause without any plural morphology and is treated as an invariable noun, which, as will be shown in Section 4, also complies with the rules of plural marking in homeland Piedmontese. The second type of minimal insertion consists of cases where inserted nouns are inflected with morphemes of the matrix language, as is the case in (2), as well as in (7), where the Spanish noun *obra* ‘facility’ takes the Piedmontese plural marker *-e* and is inflected as a Piedmontese noun.

- (7) *l’=ha fait tant-e OBR-e*
 CLIT=have:AUX.3SG do:PP many-F.PL facility(F)-PL
 ‘He built several facilities’

Before delving into the distribution of these code-mixing strategies in our data, we will first briefly present HP in more detail, since descriptions in English are still rather scarce.

3 On Heritage Piedmontese

3.1 On Piedmontese

As stated in Section 1, Piedmontese (ISO 639-3 pms), is a Gallo-Italic Italo-Romance variety predominantly spoken as a minority language within the Piedmont region, in the northwest of Italy. While it is not fully recognised as a minority language by Italian law, several bottom-up actions have been taken for its protection and revitalisation, especially by local scholars and private institutions. Not unlike most Italo-Romance varieties, Piedmontese was the main language of oral communication in Piedmont until the first half of the 20th century, when Standard Italian started spreading across the population, in schools and through other widespread means of communication. Currently,

Eugenio Gorla & Kate Bellamy

Piedmontese is endangered, due to language shift towards Italian; it is still spoken especially in rural areas of Piedmont, almost exclusively in informal and familiar domains and alongside (or in competition with) Italian. As discussed by [Regis \(2011, 2013\)](#), [Regis & Rivoira \(2023\)](#), among others, there is considerable sub-regional variation in spoken Piedmontese, but the dialect from the capital city of Piedmont, Turin, enjoys particular prestige and has provided a basis for (partial) standardisation, as well as for the elaboration of a written norm ([Brero & Bertodatti 1988](#)).

Piedmontese migration to Argentina began during the economic crisis that followed the unification of Italy in 1861. Analysing historical demographic accounts, historians (see [Nascimbene 1987](#)) identify a north-western phase in Italian migration to Argentina, approximately between the last decades of the 19th century and the outbreak of the First World War in 1914. In this period, the majority of emigrants arriving in Argentina were Piedmontese agricultural workers. The arrival of workforce from Italy, and Europe in general, was encouraged by the Argentinian government, as a means to reinforce the agricultural exploitation of the grasslands in the provinces of Córdoba and Santa Fe, which soon were named *la pampa gringa* (lit. ‘the prairie of the gringos’⁶). As emerges from [Girolitto’s](#) (2010) account, the typical settlement in this period was represented by a system of newly founded agricultural colonies, mostly inhabited by Piedmontese immigrants. Due to the distance and relative isolation from larger urban settlements, as well as to local self-administered schools, Piedmontese continued being spoken in these communities, and was even passed on to migrants from other regions or other European countries. Thus, language shift towards Spanish occurred at a much slower pace compared to other communities of Italian origin (see [Turchetta 2005](#) for an overview). Moreover, a movement for the conservation and revival of Piedmontese in Argentina started in the 1970s, in the wake of similar movements arising in Piedmont. Grammars and dictionaries of Piedmontese for Spanish speakers started to be produced (see [Gorla 2015](#) for an overview) and contacts were sought between Argentinian and Italian Piedmontese associations, often in the form of town-twinings and cultural exchanges.

Therefore, while shift towards Spanish also eventually occurred in these communities, it has been counterbalanced by an intensive revival of Piedmontese culture that has generated new (or renewed) interest in the language among Argentinians of Piedmontese origin. Thus the conditions have been created for Piedmontese to still be spoken, even if this often occurs in the context of cultural revival and activities that are programmatically dedicated to the use of the heritage

⁶*Gringo* is a term with various connotations in Latin American countries; in Argentina it normally refers to a European immigrant.

2 Code-mixing and language change

language, rather than in naturalistic daily interactions.⁷ This point seems to be of particular importance, as some definitions of heritage language – and notably Rothman (2009) – insist with great emphasis on the fact that HLs are acquired spontaneously within a naturalistic setting. In this respect, we chose to classify Piedmontese in Argentina as a heritage language *stricto sensu* for the following reasons. While it is true that language revitalisation plays an important role in shaping the observed contexts of use and linguistic practices, it must be stressed that HP was not (re)born out of revitalisation, nor is it the case that Piedmontese is acquired only through formal teaching. On the contrary, most of the existing courses are given by non-professional teachers and are aimed at creating new occasions for language use rather than formal teaching of Piedmontese grammar. Thus, based on the analysis of linguistic autobiographies (Goria 2023), we may identify different acquisitional trajectories within this community: some closer to Rothman’s definition of heritage speaker, where speakers claim to have actually learned Piedmontese within the familiar environment, and some other less typical, where knowledge of Piedmontese seems to be more dependant on formal teaching (see, e.g., Kondo-Brown 2001, for a definition of heritage language less focused on ‘native-like’ acquisition). To conclude, if we consider the bigger picture from which this data has been collected, we consider that the distinction between individuals generally termed heritage speakers and ‘learners with heritage motivation’ (Carreira & Kagan 2011) should be a matter of degree rather than a categorical one. The extent to which the multiplicity of linguistic profiles present in the community may give rise to different linguistic behaviours was not taken into account in this study and would need to be investigated in detail in future studies.

3.2 The documentation of Heritage Piedmontese

The PILAR - *Piedmontese Language in Argentina* project comprises a sociolinguistic investigation of HP. Its main aim is to compile the first spoken corpus of HP, in order to facilitate linguistic documentation. On the basis of this corpus, two main research strands emerge: (i) grammatical description of HP as a variety in

⁷An estimate of the number of speakers of heritage Piedmontese is a rather difficult task for precisely this reason: it is not clear whether the count should be limited to those who have kept some knowledge of Piedmontese from their early childhood, or if we may include also those individuals who participate in activities related to Piedmontese culture, but have only very partial knowledge of the language. Our estimate is based on the number of members of the existing Piedmontese associations, which amounts to some hundreds; we may therefore assume that actual speakers of Piedmontese will be in the same order of magnitude.

Eugenio Gorla & Kate Bellamy

its own right; and (ii) ethnographic description of the Piedmontese cultural revival taking place in Argentina. Therefore, the documentation seeks to include, also through the use of video and photography, the actual cultural practices in which HP is involved.

So far, two fieldwork trips have been conducted, one in 2019 and one in 2022, leading to the collection of around 50 hours of recordings. Informants have been recruited through a network of associations called *Federación de las Asociaciones Piemontesas de la Argentina* (Federation of Piedmontese Associations in Argentina), which coordinates the activities of various local associations of Piedmontese descendants. The research concentrates almost exclusively on small and medium-sized cities in the provinces of Córdoba and Santa Fe, where most of the heritage speakers are located. Two main types of interaction have been recorded:

- Semi-structured interviews in HP were collected by an interviewer who is a native speaker of homeland Piedmontese. Their purpose was to elicit a considerable amount of data about the grammar of HP, in a rather monological style. At the same time, parts of the recordings were conducted as group interviews involving several speakers, which often led to more informal exchanges.
- Participant observation of cultural activities focussed on HP language and culture was carried out in order to document the actual practices in which Piedmontese is involved. These include folk songs, choir singing, amateur theatre, and events of traditional cuisine.

3.3 Code-mixing in Heritage Piedmontese

It is important to stress that in the data collected in the framework of the PILAR project, HP occurs inherently embedded in multilingual practices. While on the one hand relying on local associations made it possible to get in touch with actual heritage speakers of Piedmontese who were able to engage in conversation with the researcher using the heritage language, on the other hand spontaneous uses of the language appeared restricted to a number of cultural activities specifically aimed at the celebration of Piedmontese ancestry and the preservation of Piedmontese language and culture. Based on this consideration, before discussing in detail any specific pattern of code-mixing, we need to introduce some general features of the dataset in terms of language use and bilingual practices.

Now, based on the distinction established in Section 2 between code-switching (i.e. a pragmatically meaningful alternation between multiple codes) and code-mixing (i.e. the pragmatically devoid use of lexical items of variable size within a

2 *Code-mixing and language change*

single communicative exchange), the data collected for PILAR display very few instances of conversational code-switching. Interviews were carried out in a preferably monolingual mode: either Piedmontese or Spanish were selected as the language of interaction and Spanish was used only when speakers declared themselves unable to speak Piedmontese. In other words, as is the case with several minority language contexts, the production observed here is characterised by ‘intended monolingualism’ (Dal Negro 2013, Clyne 2003): both speakers and researchers cooperate to and agree on maximising the production in the minority language, while larger switches to the dominant language tend to be avoided. It is important to note that, perhaps counter-intuitively, the choice of a monolingual mode does not rule out the possibility for code-mixing to take place, especially when it constitutes an established practice within the community. Thus, in our data code-mixing occurs consistently throughout the corpus, very often in the form of insertion of Spanish words and multi-word collocations into an otherwise HP clause. This type of behaviour, which obviously allows for some variation across speakers and contexts, probably reflects greater availability of Spanish lexicon and higher frequency of usage of such terms. However, it is relevant for the analysis that members of the community identify the use of Spanish lexicon as a distinctive feature of HP, which is sometimes given folk names such as *piemonteis merican* (‘American Piedmontese’). While some speakers, often involved in language revitalisation activities, retain purist views towards the use of Piedmontese, and consider the use of Spanish lexicon as something that must be avoided, the majority of the informants have positive or neutral attitudes towards the practice of code-mixing, which is perceived, on the contrary, as a feature in which Argentinian-Piedmontese speakers tend to identify (Goria 2023).

4 *Aims, data and methods*

The main aim of this paper is to provide the first corpus-based account of code-mixing in HP, with a specific focus on noun insertion. In fact, while in a preliminary analysis (Goria 2021) other lexical categories involved in code-mixing (e.g. verbs, discourse markers) generally showed more uniform behaviour, noun insertion displays more variation in terms of the types of code-mixing patterns. This pattern seems to be common to other situations: for example Benmamoun et al. (2013) summarise a number of previous studies on Russian, Hungarian and Hindi, where it is shown that heritage speakers of these languages tend to have more non-target productions, that is more deviations from the homeland variety, in nominal than in verbal morphology. As we will show, HP appears to comply with this general picture.

Eugenio Gorla & Kate Bellamy

As discussed in Section 1, Spanish nouns may occur in HP through (i) alternation, as (ii) maximal or (iii) minimal insertions; moreover, minimal insertions can have the form of either (iiia) bare stems without an overt morphological marking, or (iiib) Spanish stems inflected with Piedmontese morphology. Thus, we will quantitatively evaluate the distribution of these patterns. But what factors best explain the variation observed in the data? Our hypothesis is that the typology of the two contact languages, and namely the patterns of gender and number marking in the noun phrase of the respective language, may indeed influence the type of insertion observed in code-mixing. We will now review the relevant aspects of nominal morphology in the two languages involved in this study.

4.1 Spanish noun classes

In the singular, Spanish nouns are unmarked for gender and number, and these categories emerge only in agreement. However, as is typical in most Romance language, stems ending in *-o* most typically have masculine gender, and stems ending in *-a* are typically associated with feminine gender. There are two main patterns for plural formation: one for nouns ending in vowels, and one for nouns ending in consonants. These two groups take *-s* and *-es* respectively for plural marking, with very few exceptions. An overview of these patterns, for both masculine and feminine nouns, can be found in Table 2.

As can be seen from Table 2, number is marked morphologically only in the plural, while singular nouns correspond to bare stems. However, it is important to stress that some European varieties and most South American varieties of Spanish have a perspicuous tendency to phonetically reduce *-s* in syllable coda and in word-final position (Terrell 1978, Lipski 1986, Brown & Torres Cacoullos 2002). This yields realisations of the *-s* plural morpheme that include aspiration of /s/ into [h], as well as full reduction (i.e. complete loss). Morphologically the result is that the plural morpheme becomes perceptively less recognisable in vowel stems, and number in some cases can be identified only through agreement patterns, for example determiner-noun agreement makes it possible to disambiguate *el amigo* [ela'miyo] (singular) vs *los amigos* [loha'miyo] (plural).

4.2 Piedmontese noun classes

Piedmontese⁸ nouns can be divided into three main groups (see Regis & Rivoira 2023 for a recent overview). The first group, which is also the largest, includes

⁸In this overview of the grammatical features that are relevant for our interpretation of code-mixing, with 'Piedmontese' we refer to homeland Piedmontese.

2 Code-mixing and language change

Table 2: Noun classes in Spanish

	Vowel stems (M and F)	Consonant stems (M and F)
SG	<i>abrigo</i> coat(M).SG <i>mano</i> hand(F).SG <i>luna</i> moon(F).SG <i>cura</i> priest(M).SG <i>padre</i> father(M).SG <i>madre</i> mother(F).SG	<i>inquietud</i> concern(F).SG <i>nivel</i> level level(M).SG <i>Inglés</i> English(M).SG
PL	<i>abrigo-s</i> coat(M)-PL <i>mano-s</i> hand(F)-PL <i>luna-s</i> moon(F)-PL <i>cura-s</i> priest(M)-PL <i>padre-s</i> father(M)-PL <i>madre-s</i> mother(F)-PL	<i>inquietud-es</i> concern(F)-PL <i>nivel-es</i> level(M)-PL <i>Inglés-es</i> English(M)-PL

invariable nouns ending in vowel or in consonant, both masculine and feminine. The second group includes feminine nouns ending in *-a* in the singular and in *-e* in the plural. The third group includes a small group of masculine nouns ending in *-l* in the singular and in *-i* in the plural.⁹

As illustrated in Table 3, the largest group of Piedmontese nouns are invariable for number, with the biggest exception being feminine nouns belonging to the *a/-e* group and, to a lesser extent, by the synchronically unproductive *-l/-i* group.

⁹As pointed out by Ricca (2008) a new class appears to be developing in Piedmontese, whereby nouns borrowed from Italian retain their original inflection, as is the case for the word *chilometro* ‘kilometre’, for example, which occurs both as an invariable and with the Italian plural inflection *chilometri*. However, so far there have not been any quantitative accounts of this phenomenon and, also given its minor relevance to the issues presented here, we will not discuss it further in this paper.

Eugenio Gorla & Kate Bellamy

Table 3: Noun classes in Piedmontese

	Invariables in consonant (M and F)	Invariables in vowel (M and F)	Feminine nouns in -a/-e	Masculine nouns in -l/-i
SG	<i>amis</i> friend(M):SG; <i>man hand</i> (F):SG	<i>mare</i> mother(F):SG <i>pare</i> father(M):SG	<i>crav-a</i> goat(F)-SG	<i>caval</i> horse(M):SG
PL	<i>amis</i> friend(M):PL <i>man hand</i> (F):PL	<i>mare</i> mother(F):PL <i>pare</i> father(M):PL	<i>crav-e</i> goat(F)-PL	<i>cavai</i> horse(M):PL

Gender follows the typical pattern of Romance languages, and is determined by a combination of formal and semantic parameters. Simply put, nouns belonging to the *-a/-e* class are formally marked as feminine; in all the other cases, it is impossible to determine the gender of a noun according to formal criteria.

In contrast, gender and number are identifiable through agreement patterns within the NP, as articles and the other determiners have separate endings for all genders and numbers. See Tables 4-6 for a full overview.

Table 4: Paradigm of Piedmontese definite articles

Definite article	M	F
SG	ël /əɫ/	la /la/
PL	ij /i/	le /le/

Table 5: Paradigm of Piedmontese indefinite articles

Indefinite article	M	F
SG	<i>un</i> /yŋ/	<i>na</i> /na/
PL	<i>dij</i> /dij/	<i>ëd le</i> /əd le/

2 Code-mixing and language change

Table 6: Paradigm of Piedmontese quantifier *tut* ‘all’

Quantifier <i>tut</i> ‘all’	M	F
SG	<i>tut</i> /tyt/	<i>tut-a</i> /tyta/
PL	<i>tut-i</i> /tyti/	<i>tut-e</i> /tyte/

It should be noted, however, that the situation presented in this section most accurately describes the partially standardised variety, whose model is the dialect of the capital city of Piedmont: Turin (Italian: Torino). Partially different patterns may be found in specific areas (see [Berruto 1974](#), [Telmon 2001](#), [Regis & Rivoira 2023](#)).

4.3 Comparison between the two systems and predictions

Gender assignment works in a similar way in Spanish and Piedmontese: nouns denoting human entities are treated as masculine or feminine based on biological sex, while non-human nouns are as masculine or feminine depending on their noun class. Differences may be found, though, in terms of marking: as shown in 4.1 and 4.2, masculine gender in Spanish tends to be associated with the *-o/-os* class, and feminine gender with the *-a/-as* class, whereas in Piedmontese we find such a clear association only for the class of *-a/-e* feminines. This yields an asymmetrical situation, which may pose a problem in a language contact situation: while a structural equivalence may be identified in Spanish and Piedmontese for feminine nouns, due to similarity in grammatical marking, the same does not hold for masculine nouns, where the two languages differ.¹⁰

With regard to number marking, in Piedmontese, this varies based on the noun class: while the *-a/-e* class of feminine nouns overtly marks number (at least in the plural), all the other nouns are invariable, except for those ending in *-l*. In contrast, Spanish number is formally marked only in the plural, but it is consistently marked on all noun classes with perhaps a few exceptions. In other words, Spanish and Piedmontese nouns are formally equivalent in the singular, but show differences in the plural.

¹⁰An anonymous reviewer suggests that gender mismatches, namely, nouns that are masculine in one language and feminine in the other, should be addressed in this study. This was not possible, however, since in the subset of recordings that was analysed for this study, we had no evidence of gender mismatches.

Eugenio Gorla & Kate Bellamy

While in general the choice between maximal and minimal insertion is less predictable based on typological factors, and tends to be associated with usage-based factors (see [Hakimov 2021](#)), we are able to make some predictions about the choice of what type of minimal insertion will be observed, according to the morphosyntactic treatment of nouns in the two languages involved. Thus, it follows from the previous description that in Piedmontese-Spanish code-mixing less variation is to be expected in the singular, where the marking strategies are equivalent, than in the plural, where the two languages differ. More specifically, we expect the bare noun strategy of minimal insertion to be preferred, as it does not violate the constraints of Piedmontese morphology. Moreover, we must also remember that since in colloquial Spanish syllable-final -s (which includes the plural marker -s) tends to undergo phonetic reduction, nouns that show this feature are *de facto* invariables and also a locus for potential structural equivalence. We must furthermore expect a different treatment of masculine and feminine nouns, since the two are marked differently in Piedmontese: while the bare noun strategy will probably be preferred for masculine nouns, feminine nouns may either occur as bare stems, or be treated as Piedmontese nouns of the *a/-e* class.

5 Data

For the present study, a sample of five hours was selected from the corpus, involving 22 speakers of different ages, ranging from 45 to 90. With only two exceptions, the participants were born in Argentina and, based on their linguistic biographies¹¹, we were able to categorise them as fully-fledged heritage speakers who acquired Piedmontese within the household in their childhood and subsequently stopped using the language except for a few highly specific cultural practices connected with the preservation of Piedmontese language and culture. The two speakers who were born in Italy can be assimilated to this category as they arrived in Argentina in their childhood and attended Spanish-speaking schools; moreover, one informant was also the child of a transnational family and even though she was born in Italy, her father was already working in Argentina. See

¹¹Linguistic biographies were recorded during a semi-structured interview, aimed both at collecting relevant metadata for each participant, and at inviting the production of oral narratives. The questions concerned the time and place of arrival of the family in Argentina, family language policies concerning the use of Spanish and Piedmontese, and actions taken by single individuals for learning or improving their knowledge in the heritage language. Examples of such actions are participating to the activities of the local association, taking courses of Piedmontese, or travelling to Piedmont.

2 Code-mixing and language change

Table 7 for details.¹²

After having transcribed the spoken data in ELAN (Wittenburg et al. 2006, Sloetjes & Wittenburg 2008), we annotated the gender and number of the noun and the types of code mixing observed for all the Spanish nouns occurring in the data.

6 Results

The analysis of the selected recordings yielded a total of 318 Spanish nouns, of which 16 were classified as instances of alternation, 66 as maximal insertions, and 236 as minimal insertions¹³. Table 8 summarises these results.

As far as NPs are concerned, we find rather uniform behaviour across the community, as insertional strategies represent around 95% of all the switches considered for this study. Minimal insertions in particular are the most represented type of code-mixing. Among minimal insertions, we found the majority (or 64% of all types of code-mixes in the corpus) to be instances of bare noun insertion, that is, cases where Spanish nouns ending in a vowel (8) or in a consonant (9) are treated as Piedmontese invariable nouns.

(8) a j'è ëd TAMBO gròss
 there_is ART.IND.PL dairy_farm(PL) large
 'There are large dairy farms'

(9) ël pì cit dij VARÓN
 the:M.SG more young of.the:M.PL male
 'The youngest of the males'

Minimal insertions, where Spanish nouns take Piedmontese morphological gender and number marking, represent 10% of the corpus and include a variety of qualitatively different cases. First of all we have cases of feminine plural

¹²The table contains the actual initials of the informants. All the speakers voluntarily participated in this study, and agreed for the data to be shared, presented in publications, and not be anonymised. Recordings were made after introducing the aims of the research and asking explicit permission to record. In 2022 the collection was organised in cooperation with the national Piedmontese association, who took care of informing all the participants about the recordings. Either a recorded consent or a signed one, depending on the occasion, was asked to the participants.

¹³The annotated data can be found at the following link: https://osf.io/mu6gd/?view_only=54084412706e4d018398c1b6db566f09

Eugenio Gorla & Kate Bellamy

Table 7: Participants in the present study

Spkr	Birth	Born in	Gender	Interaction	City	Recorded
JB	1956	Argentina	M	Interview	Freyre (Córdoba)	2019
AB	1933	Argentina	M	Interview	Arroyito (Córdoba)	2019
GI	na	Argentina	F	Interview	Arroyito (Córdoba)	2019
LA	Na	Argentina	M	Interview	Arroyito (Córdoba)	2019
EM	1939	Argentina	M	Interview	Córdoba	2019
EO	1945	Argentina	F	Interview	Devoto (Córdoba)	2019
AS	1978	Argentina	M	Group Interview	Freyre (Córdoba)	2019
AM	1978	Argentina	M	Group Interview	Zenon Pereyra (Santa Fe)	2019
LM	1943	Italy	M	Group Interview	Zenon Pereyra (Santa Fe)	2019
RC	1940	Argentina	M	Interview + Informal conversation	Brinkmann (Córdoba)	2022
ND	na	Argentina	M	Informal conversation	Brinkmann (Córdoba)	2022
AG	1971	Argentina	F	Group Interview	Las Varillas (Córdoba)	2022
JG	1941	Argentina	M	Group Interview	Las Varillas (Córdoba)	2022
NL	1966	Argentina	M	Group Interview	Las Varillas (Córdoba)	2022
OC	1952	Argentina	F	Group Interview	Las Varillas (Córdoba)	2022
GB	1964	Argentina	F	Group Interview	Las Varillas (Córdoba)	2022
VC	1944	Italy	F	Group Interview	Córdoba	2022
HG	1960	Argentina	M	Group Interview	Córdoba	2022
RR	1937	Argentina	M	Group Interview	Córdoba	2022
MB	1938	Argentina	M	Group Interview	Córdoba	2022
SB	1940	Italy	M	Group Interview	Córdoba	2022
UB	1935	Argentina	M	Group Interview	Córdoba	2022

2 Code-mixing and language change

Table 8: Frequencies observed for each type of code-mixing in HP-Spanish

CM Type		N	%
Alternation		16	5
Maximal insertion		63	20
Minimal insertion	Bare nouns	205	65
	Piedmontese morphology	33	10
TOTAL		318	100

nouns, which are normally inflected like Piedmontese nouns belonging to the *-a/-e* class; the difference is obviously only noticeable in the plural, as the singular is homophonous; consider example (10).

- (10) l’=ha fait tant-e OBR-e
CLIT=have:AUX.3SG do:PP many-F.PL facility(F)-PL
‘He built several facilities’

In other cases, in code-mixing contexts Spanish nouns receive inflectional marking in Piedmontese which might reveal partial reanalysis; consider example (11).

- (11) ël sigilin ai dis-o ël BALD-O
the:M.SG bucket 3SG:OBL say-3PL the:M.SG bucket-M.SG
‘The bucket, they call it *baldo*’

In this case, the Spanish word *balde* displays some form of morphological integration through the introduction of the ending /o/, realised as [u] or [o]. While the data are too scarce to provide a detailed analysis, we may hypothesise that at some point a correspondence rule (Thomason 2007) was created, so that Spanish bare stems in *-o* were produced with a final [u] complying with a rule of Piedmontese phonology whereby etymological unstressed /o/ is realised as [u]. It could also be the case that this final [u] was later also extended to Spanish nouns ending with other vowels.¹⁴ A similar phenomenon occurs with some plural nouns that take the marker *-i*, as in example (12).

¹⁴ A piece of evidence in favour of this interpretation is that most examples of this type come from metalinguistic examples where speakers report on typical linguistic habits of the community.

- (12) j era tut-i MONT-i li
CLIT be:IMPF.3SG all-M.PL bush-PL there
'It was all bush there'

Here the Spanish noun *monte* is inflected in the plural with the ending *-i*: as shown in Section 4.2, this ending is absent in Piedmontese nominal inflection, but is present in determiners, pronouns and in some Italian borrowings. As only some speakers are also speakers of Italian, it is not immediately possible to identify the origin of this phenomenon; we argue however that it should be considered as an innovation typical of this variety. As emerges from this preliminary qualitative account of insertion in HP, gender and number play an important role in determining the observed code-mixing patterns; therefore, we analysed how the data vary according to gender and number, obtaining the distribution presented in Table 9.

Table 9: Frequencies of each pattern of switching according to gender and number of the switched noun

	M.SG	M.PL	F.SG	F.PL
Maximal insertion	24 (19.2%)	7 (12.73%)	27 (27%)	5 (23.81%)
Bare nouns	95 (76%)	33 (60 %)	73 (73%)	4 (19.05%)
Inflected nouns	6 (4.80%)	15 (27.27%)	0	12 (57.14%)
Total	125	55	100	21

From the analysis of this distribution, it emerges that singular marking is more regular in code-mixing than plural marking, with the bare-stem type described above predominating. In the plural, where the two contact languages show typologically different features, more strategies are found. Furthermore, differences are also found between masculine and feminine nouns: the first, in most cases, are treated as Piedmontese masculines in bilingual discourse, that is, typically invariable; the latter, being treated like Piedmontese feminines, display a higher frequency of cases with Piedmontese inflection.

We used Fisher's exact test on the data from Table 9, after excluding the maximal insertions ($N=236$ after exclusion), in order to test whether the distribution of the two types of minimal insertion can be dependent on the gender or number of the switched noun. The distribution transpired to be non-significant for gender, but significant for number ($N=236$, $p<0.00001$), thus confirming that nouns

2 Code-mixing and language change

inflected with Piedmontese morphology are found more frequently in the plural than in the singular. See Figure 1. Indeed, the only singular cases of nouns with overt inflection in the singular are cases like *mat-o* discussed above.

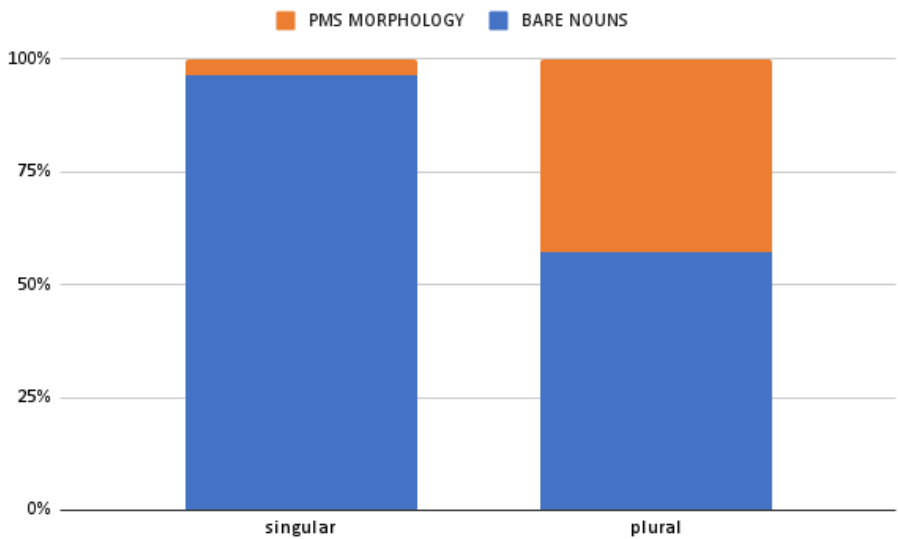


Figure 1: Ratios of bare stems vs inflected nouns in the singular and plural

The picture becomes even clearer if we focus exclusively on plural nouns, which constitute the area where most variation is observed. Based on the distribution of Table 9, we thus repeated Fisher’s exact test only on plural nouns in order to test the distribution for statistical significance. The test returned a *p value* of 0.002, thus demonstrating that Spanish masculine nouns occur more frequently as bare stems in code-mixing, and are therefore treated as invariables, whereas feminine nouns are more frequently matched to the Piedmontese *-a/-e* class, and thus are inflected with the Piedmontese plural marker *-e*; see Figure 2.

7 Discussion and conclusions

The analysis of Spanish-Piedmontese code-mixing carried out in this study enabled us first of all to publish new data in English about a language, Piedmontese, that is endangered in its homeland and that is spoken in an even more fragile ecology in the post-migration setting of Argentina described in the previous sections.

Eugenio Gorla & Kate Bellamy

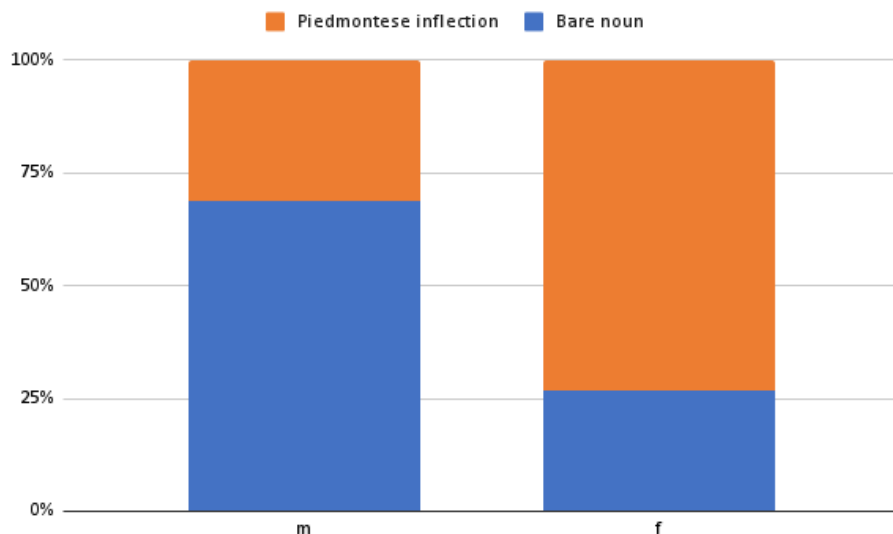


Figure 2: Ratios of the two strategies of plural marking observed for masculine and feminine nouns (N=63)

We therefore regard the present contribution as a milestone in the documentation of Heritage Piedmontese.

We were able to identify a number of tendencies that characterise the bilingual situation presented in this study, with the main aim of determining what type of code-mixing predominated among and between the speakers. To this end, the data revealed a much greater preference for insertion over alternation, which appears to be a marginal phenomenon in the corpus. This preference can be interpreted as a tendency towards what *Dal Negro (2013)* (see also *Clyne 2003*) refers to as “intended monolingualism”, that is the choice to establish the heritage language as the base language of the interaction, and as the matrix language of the clause, thereby avoiding larger shifts towards Spanish, as is the case with alternation. As mentioned in Section 3, the reason for this choice is probably an effect of the fact that speakers generally show positive attitudes towards HP and tend to position themselves as proficient speakers of the language, and therefore they tend to display their knowledge also by avoiding major switches into Spanish. Other evidence in favour of this view is given by the presence of self-corrections and feedback towards other speakers when too much Spanish was used in the conversation. At the same time, one of the most salient features in HP is indeed the abundant presence of Spanish lexicon that characterises code-mixing (as op-

2 Code-mixing and language change

posed to code-switching; see the discussion in 3.3): most of the evaluations collected on the field identify this phenomenon as one of the hallmarks of *piemonteis merican*, that is the name given by the community to HP.

In contrast, a possible explanation for the observed distribution resides in the methodology used for data collection: since the data were collected with the primary purpose of documenting HP, the interviewer systematically used that language during data collection (see also the methodological remarks in [Blommaert & Dong 2010](#)), thus inviting monolingual contributions in this language, and possibly preventing other conversational styles where alternation between Spanish and HP occurs more frequently.

Quantitative analysis of noun insertion showed that maximal insertions represent a marginal phenomenon in the corpus compared to minimal insertions, which are the most frequent pattern. We then moved on to evaluate if, based on the typology of gender and number marking on nouns in the two contact languages, it was possible to predict the observed type of switch. The results seem to confirm our hypothesis: as a general consideration, we may conclude that at the points where the two languages are structurally equivalent, the same structure is maintained in code-mixing; and thus, although from a partially different theoretical perspective, we may observe that [Poplack's](#) (1980) equivalence principle is well reflected in this data. The prevalence of bare nouns in the singular is best interpreted as a reflection of the fact that in both Spanish and Piedmontese singular number is zero-marked; in fact, most of the singular nouns that have overt Piedmontese morphemes are derived words such as *chacar-é* 'farmer' (Spanish *chacarero*), and *alambr-à* 'wiring' (Spanish *alambrado*), which obtain singular marking through the adoption of a Piedmontese derivational suffix.

Plural nouns, on the other hand, displayed more variation in code-mixing, with significant differences between masculine and feminine nouns. While the bare noun strategy is preferred for masculine nouns, feminines more frequently adopt Piedmontese overt morphology. Again, the Piedmontese number system provides an explanation for this difference, as masculine nouns are mostly invariables, whereas feminine nouns have overt inflection. Furthermore, this tendency may also be reinforced by a peculiarity of spoken Argentinian Spanish, as the plural marker *-s* is involved in the phenomenon of phonetic reduction of syllable final *-s*, with the effect of losing plural marking.

At the same time, we were able to identify a small number of cases of insertions that are not immediately predictable from the grammar of the two contact languages. We saw that some Spanish nouns are inserted into Piedmontese with a final /o/, both in the singular and in the plural, and in some cases, Spanish nouns in *-e* take *-i* as a plural marker in code-mixing. While the data are too scarce to

Eugenio Gorla & Kate Bellamy

make a more concrete claim, such instances might represent innovations that have arisen in the heritage language scenario and which have, at some point, diffused into the community, as they also appear in the first author's field notes and in recordings that were not considered for this study. To conclude, this research supports the notion that identifying typological similarities and differences in the contact languages may indeed help us predict the type of code-mixing that will be observed. Thus, on the one hand the greater propensity for bare nouns seems to be determined by the fact that Spanish nouns are adapted to a grammatical frame where masculine nouns are invariables. Furthermore, we saw that if we take into account spoken Spanish, the equivalence is even greater due to phonetic reduction. Conversely, structural equivalence is also responsible for the presence of overtly inflected plural feminines, because a clearer match is found between Spanish and Piedmontese feminine nouns.

As a methodological note, it must also be observed that in this study we used as a benchmark for evaluating structural innovations in HP the written variety of Torino, which has provided a basis for grassroots revitalisation initiatives within the community (in Italy) and for grammatical description. This might be problematic in principle, as it could blur out internal variation in the input, which is – as is well known – one of the sources of language change in heritage languages. However, we must stress that none of the non-target forms that were identified in the dataset can be associated with any of the local dialects of Piedmontese. Internal variation in Piedmontese will need to be necessarily taken into account if this research is extended to other sectors of grammar where greater variation is observed (e.g. determiners).

Finally, the observation of this situation at the synchronic level of code-mixing also leads us to some observations about diachronic language change (see Backus 2005). While we are not able to reconstruct how code-mixing was in the past and what the origins of this situation might be, we may ask whether the observed preference for bare nouns, at least in the masculine, may represent an emergent change in HP, especially on account of the observed overextensions such as Spanish *balde* 'bucket' being realised as *baldo*. Also the presence of emic identification of these forms as typical of HP seems to point in this direction. Such a hypothesis of course needs to be tested on a much larger dataset, also taking into consideration other factors such as internal variation in the community, the type of literacy that the speakers have in the heritage language, and so on. However, if supported, it could provide relevant evidence for the relation between code-mixing at the synchronic level and diachronic language change.

References

- Aalberse, Suzanne, Ad Backus & Pieter Muysken. 2019. *Heritage languages: A language contact approach* (Studies in Bilingualism 58). Amsterdam: John Benjamins. DOI: [10.1075/sibil.58](https://doi.org/10.1075/sibil.58).
- Auer, Peter. 1999. From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *International Journal of Bilingualism* 3(4). 309–332. DOI: [10.1177/13670069990030040101](https://doi.org/10.1177/13670069990030040101).
- Auer, Peter. 2005. Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations. In Nicole Delbecque, Johan van der Auwera & Dirk Geeraerts (eds.), *Perspectives on variation* (Trends in Linguistics. Studies and Monographs 163), 7–42. Berlin: De Gruyter Mouton. DOI: [10.1515/9783110909579.7](https://doi.org/10.1515/9783110909579.7).
- Auer, Peter. 2014. Language mixing and language fusion: When bilingual talk becomes monolingual. In Juliane Besters-Dilger, Cynthia Dermarkar, Stefan Pfänder & Achim Rabus (eds.), *Congruence in contact-induced language change*, 294–334. Berlin: De Gruyter. DOI: [10.1515/9783110338454.294](https://doi.org/10.1515/9783110338454.294).
- Backus, Ad. 2005. Codeswitching and language change: One thing leads to another? *International Journal of Bilingualism* 9(3–4). 307–340. DOI: [10.1177/13670069050090030101](https://doi.org/10.1177/13670069050090030101).
- Benmamoun, Elabbas, Silvina Montrul & Maria Polinsky. 2013. Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics* 39(3–4). 129–181. DOI: [10.1515/tl-2013-0009](https://doi.org/10.1515/tl-2013-0009).
- Berruto, Gaetano. 1974. *Piemonte e Valle D'Aosta* (Profilo dei dialetti italiani 5). Pisa: Pacini.
- Berruto, Gaetano. 2012. *Sociolinguistica dell'italiano contemporaneo*. 2nd edn. (Manuali universitari 131). Rome: Carocci.
- Blommaert, Jan & Jie Dong. 2010. *Ethnographic fieldwork: A beginner's guide*. 1st edn. Bristol: Multilingual Matters. DOI: [10.21832/9781847692962](https://doi.org/10.21832/9781847692962).
- Brero, Camillo & Remo Bertodatti. 1988. *Grammatica della lingua piemontese*. Torino: Piemont/Europa.
- Brown, Esther & Rena Torres Cacoullos. 2002. ¿Qué le vamoh aher?: Taking the syllable out of Spanish /s/ reduction. *University of Pennsylvania Working Papers in Linguistics* 8(3). Papers from NWAV 30, 17–31. <https://repository.upenn.edu/handle/20.500.14332/45605>.
- Carreira, Maria & Olga Kagan. 2011. The results of the National Heritage Language Survey: Implications for teaching, curriculum design, and professional development: The results of the National Heritage Language Survey. *Foreign Language Annals* 44(1). 40–64. DOI: [10.1111/j.1944-9720.2010.01118.x](https://doi.org/10.1111/j.1944-9720.2010.01118.x).

Eugenio Gorla & Kate Bellamy

- Cerruti, Massimo, Riccardo Regis & Eugenio Gorla. 2025. Baselines in contact. The case of piedmontese in Argentina. In Eugenio Gorla & Margherita Di Salvo (eds.), *Italo-Romance heritage languages. Multiple approaches*, 172–198. Amsterdam: Benjamins.
- Clyne, Michael. 2003. *Dynamics of language contact: English and immigrant languages*. 1st edn. Cambridge: Cambridge University Press. DOI: [10.1017 / CBO9780511606526](https://doi.org/10.1017/CBO9780511606526).
- Dal Negro, Silvia. 2013. Dealing with bilingual corpora: Parts of speech distribution and bilingual patterns. *Revue française de linguistique appliquée* 18(2). 15–28.
- De Mauro, Tullio. 1963. *Storia linguistica dell'Italia unita*. Bari: Laterza.
- Deuchar, Margaret, Pieter Muysken & Sung-Lan Wang. 2007. Structured variation in codeswitching: Towards an empirically based typology of bilingual speech patterns. *International Journal of Bilingual Education and Bilingualism* 10(3). 298–340. DOI: [10.2167/beb445.0](https://doi.org/10.2167/beb445.0).
- Doglioli, Nicolas. 2022. *Connotazioni e prestigio del termine gringo all'interno della comunità piemontese in Argentina*. Torino: Università di Torino. (Doctoral dissertation).
- Giolitto, Marco. 2000. Pratiche linguistiche e rappresentazioni della comunità piemontese d'Argentina. *Éducation et Sociétés Plurilingues* 9. 13–19.
- Giolitto, Marco. 2010. *La communauté piémontaise d'Argentine: Évolution, fonction et image du piémontais dans la Pampa gringa argentine* (Forum Sprachwissenschaften 9). München: Meidenbauer.
- Giolitto, Marco. 2016. *Palabras de gringos. El uso del piamontés en la vida cotidiana de los habitantes de la Pampa Gringa*. Rosario: Prohistoria Ediciones.
- Gorla, Eugenio. 2015. Il piemontese di Argentina. Considerazioni generali e analisi di un caso. *Rivista Italiana di Dialettologia, Lingue Dialetti e Società* 39. 127–158.
- Gorla, Eugenio. 2021. Il piemontese in Argentina. Preliminari per un'analisi sociolinguistica. In Gabriele Iannaccaro & Simone Pisano (eds.), *Intrecci di parole. Esperienze di pianificazione del plurilinguismo, in Europa e fuori dell'Europa*, 233–250. Alessandria: Dell'Orso.
- Gorla, Eugenio. 2023. Il piemontese in Argentina. Aspetti linguistici ed etnografici. In Daniela Mereu & Silvia Dal Negro (eds.), *Confini nelle lingue e tra le lingue*, 219–235. Milan: Società di Linguistica Italiana.
- Gorla, Eugenio & Margherita Di Salvo. 2023. An Italo-Romance perspective on heritage languages. *Italian Journal of Linguistics* 35(1). 45–70.

2 Code-mixing and language change

- Hakimov, Nikolay. 2021. *Explaining Russian-German code-mixing: A usage-based approach* (Contact and Bilingualism 3). Berlin: Language Science Press. DOI: [10.5281/ZENODO.5589446](https://doi.org/10.5281/ZENODO.5589446).
- Kerswill, Paul. 2006. Migration and Language. Migration und Sprache. In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier & Peter Trudgill (eds.), *Sociolinguistics / Soziolinguistik*, vol. 3, 2271–2284. Berlin: De Gruyter.
- Kondo-Brown, Kimi. 2001. Heritage language instruction for post-secondary students from immigrant backgrounds. *Heritage Language Journal* 1(1). 1–25. DOI: [10.46538/hlj.1.1.1](https://doi.org/10.46538/hlj.1.1.1).
- Lipski, John M. 1986. Instability and Reduction of /s/ in the Spanish of Honduras. *Revista Canadiense de Estudios Hispánicos* 11. 27–47.
- Muysken, Pieter. 2000. *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.
- Muysken, Pieter. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition* 16(4). 709–730. DOI: [10.1017/S1366728912000727](https://doi.org/10.1017/S1366728912000727).
- Myers-Scotton, Carol. 1993. *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon.
- Myers-Scotton, Carol (ed.). 2002. *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press. DOI: [10.1093/acprof:oso/9780198299530.002.0003](https://doi.org/10.1093/acprof:oso/9780198299530.002.0003).
- Nagy, Naomi. 2017. Documenting Variation in (Endangered) Heritage Languages: How and Why? *Language Documentation & Conservation* 13. 33–64.
- Nascimbene, Mario. 1987. Storia della collettività italiana in Argentina (1835-1965). In Francis Korn (ed.), *La popolazione di origine italiana in Argentina*, 203–613. Torino: Fondazione Agnelli.
- Polinsky, Maria. 2018. *Heritage languages and their speakers*. 1st edn. Cambridge: Cambridge University Press. DOI: [10.1017/9781107252349](https://doi.org/10.1017/9781107252349).
- Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics* 18(7/8). 581–618. DOI: [10.1515/ling.1980.18.7-8.581](https://doi.org/10.1515/ling.1980.18.7-8.581).
- Regis, Riccardo. 2011. Koinè dialettale, dialetto di koinè, processi di koinizzazione. *Rivista Italiana di Dialettologia, Lingue Dialetti e Società* 35. 7–36.
- Regis, Riccardo. 2013. Può un dialetto essere standard? *Vox Romanica* 72. 151–169.
- Regis, Riccardo & Matteo Rivoira. 2023. *Dialetti d'Italia. Piemonte e Valle d'Aosta*. Rome: Carocci.
- Ricca, Davide. 2008. Tratti instabili nella sintassi del piemontese contemporaneo. In Sabine Heinemann (ed.), *Sprachwandel und (Dis-)Kontinuität in der Roma-*

Eugenio Gorla & Kate Bellamy

- nia (Linguistische Arbeiten 521), 113–128. Tübingen: Niemeyer. DOI: [10.1515/9783484970403.113](https://doi.org/10.1515/9783484970403.113).
- Rothman, Jason. 2009. Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism* 13(2). 155–163.
- Sloetjes, Han & Peter Wittenburg. 2008. Annotation by category: ELAN and ISO DCR. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis & Daniel Tapias (eds.), *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco*, 816–820. Paris: European Language Resources Association (ELRA). <https://aclanthology.org/L08-1034/>.
- Telmon, Tullio. 2001. *Piemonte e Valle d'Aosta*. 1st edn. (Profili linguistici delle regioni). Roma: Laterza.
- Terrell, Tracy. 1978. La aspiración y elisión de /s/ en el español porteño. *Anuario de Letras* 16. 41–66.
- Thomason, Sarah. 2007. Language contact and deliberate change. *Journal of Language Contact* 1(1). 41–62. DOI: [10.1163/000000007792548387](https://doi.org/10.1163/000000007792548387).
- Turchetta, Barbara. 2005. *Il mondo in italiano. Varietà e usi internazionali della lingua*. Bari: Laterza.
- Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann & Han Sloetjes. 2006. ELAN: A professional framework for multimodality research. In Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odijk & Daniel Tapias (eds.), *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy*, 1556–1559. Paris: European Language Resources Association (ELRA). <https://aclanthology.org/L06-1082/>.

Chapter 3

Verinaiss ‘very nice’: The integration of English in Estonian teenagers’ instant messages

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

This paper analyzes Estonian teenagers’ code-switching with L2 English in instant messaging (IM) language. We focus on the orthographic and morphological integration of English units and consider the variability inherent to the IM register. The analysis is based on the newly compiled Corpus of Estonian Teenagers’ Language, containing spoken and IM language from informants aged 10-18; for this study, we analyze the IM conversations, containing 60,000 words from 102 informants, 99 of whom use at least some English. On average, 7.5% of words are in English, with great fluctuation between speakers and conversations.

Our data support the gradient view of morphosyntactic integration expressed by [Deuchar \(2020\)](#). It has been shown that IM chat language deploys strategies imitating spoken and written language, as well as bearing features distinguishing it from both: code-switching in the IM register also shows features unlike either spoken or written registers, including the variability in orthographic integration. Our data show how Estonian teenagers make ample use of their L2 English in conversations with other Estonian teens, as an extension of their Estonian language usage. Orthographic integration is optional, playful, and sometimes serves to flag unintegrated phonology. Morphological integration is more fundamental to language fluency, and we find that over 80% of nouns and verbs in contexts requiring morphological marking are marked, whereas adjectives are much less likely to be marked. Despite the register’s preference for unconventional usage, code-switching in our data follows the expectations based on spoken-language code-switching more often than it flouts them.

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

1 Introduction

Teens are considered to be innovative language users and potential initiators of language change (Cheshire 2005, Leppänen 2007). Eckert's depiction of teenagers as linguistic "movers and shakers" (Eckert 1997: 52) refers to their creativity and use of linguistic resources to create new linguistic practices, identities, and values (see Stenström, Stenström et al. 2002, Nortier 2018). The continual emergence of new linguistic practices is witnessed in today's digitally mediated conversations, where the diverse linguistic, cultural, and audiovisual resources and the endless stream of "content" provides a wealth of opportunity for communicative creativity and innovation. Additionally, the online environment experienced by teens in many parts of the world includes a blend of languages, fostering POLYLINGUALISM (Jørgensen 2008) and TRANSLANGUAGING (García 2009, García & Wei 2014), both of which include the idea that multilingual language use does not require language fluency, noting that any linguistic resources can be drawn upon in a communicative context to varying effects.

While teenagers' language use has been described as a distinct CODE or REGISTER (see Androutsopoulos 2005, Tagliamonte 2016b), those teenagers inhabiting a multilingual environment raise intriguing questions for researchers regarding innovation, variation, and the use of linguistic resources for negotiating social identity. The 'translanguaging' in linguistic practices among young people in urban spaces differs from what has been traditionally examined in multilingualism and code-switching research, both in the language backgrounds of the speakers and in their language usage. This approach ascribes dynamic fluidity to the speakers, who make use of their entire linguistic repertoire and competencies to express themselves. In addition to studies which have focused on English-speaking teenagers, previous studies have described youth registers in the context of other languages including Swedish (Kotsinas 2004), Danish (Quist 2010), Dutch (Schoonen & Appel 2005), Belgian Dutch, or Flemish (Schuring & Zenner 2022), and Spanish (Jørgensen 2013). Earlier studies have mostly focused on spoken language, but researchers have emphasized the need for more research on computer-mediated communication (e.g., Quist 2010: 10).

In this paper, we investigate the language used in instant messaging (IM) by Estonian teenagers who inhabit a multilingual media environment despite the relatively monolingual character of their offline environments. In the context of today's social interaction, IM exchanged on digital devices has taken the place of much spoken conversation and are key to understanding the communicative and social behavior of contemporary youth (Tagliamonte & Denis 2008, Tagliamonte 2016a). IM conversations include a vast array of visual, textual, and intertextual

3 *The integration of English in Estonian teenagers' instant messages*

options for conveying both meaning and affect: teenagers' communication reflects a polyphonic, multimodal discourse in which they navigate seamlessly. Among these multiple layers is their use of multiple languages. The Estonian teens draw on Estonian and English, with varying amounts of Russian as well, using twice as much English when they chat online with each other as in their spoken conversations (Vihman et al. 2022; cf. Androutsopoulos 2005). As such, they embody the recent phenomenon of “networked multilinguals” (Androutsopoulos 2015), online communicators whose experience with English may be minimal offline, but who embrace a multilingual mode of communication online. How the languages are combined in this “third register” – differing from both written and spoken language – requires empirical investigation, as it sheds light on the current generation of youth residing in a mediatized and globalized communicative space. The research that has investigated multilingual youth in this context has mostly focused on migrants. Here, we focus on youth who have become multilingual in large part in and through the use of globalized media.

In this study, we investigate how Estonian adolescents use English linguistic elements – by far the most frequent non-Estonian insertions – in the context of online social interaction among Estonian-speaking friends. This indicates whether they conceive of their conversation structure as following the frame of a single language or not (e.g., Myers-Scotton 2006). We employ combined quantitative and qualitative analysis of multilingual practices in the IM conversations of Estonian teenagers, focusing on two research questions: To what extent and in what ways are English elements integrated into (primarily) Estonian conversations orthographically? And to what extent and in what ways are they integrated morphologically? The data derive from the Teen Speak in Estonia chat corpus,¹ containing 109 IM conversations between 102 unique participants aged 10 to 18.

Because Estonian spelling is phonetically transparent and hence is taken by speakers to directly indicate pronunciation, the orthographic renditions of words in the informal register of IM can be played with to indicate phonological (non-)integration of linguistic units into the matrix code. We ask whether the teenagers in our study make systematic use of orthographic integration (Estonianized spelling) to indicate phonological integration (indicating Estonianized pronunciation), and whether orthographic and morphological integration are correlated in the same way that has been found for phonological integration between English and Spanish (Bessett 2017; see also discussion in Deuchar 2020). Orthographic integration is variable, as shown in *verinaiss* ‘very nice’ compared with the unintegrated *verygood* (both attested in the corpus); variants of ‘thank

¹“Teen Speak in Estonia” project: www.sisu.ut.ee/teke (02 May, 2023).

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

you’ – e.g., *tänk ju* (see Section 4.2.1.) – can be taken to clearly evoke pronunciation as well. These questions are important both for understanding the way the participants deploy their multilingualism for constructing meaning and identity, but also for research into code-switching. The relationship between integration in morphology and phonology has long been discussed (e.g., Poplack 1980), but the relation between phonological and orthographic integration may take on a new significance in IM chats, which are able to – but not compelled to – mimic pronunciation.

The spontaneity of the IM context is expected to facilitate a range of solutions to the question of morphological integration. While various structural and social constraints on code-switching have been outlined in the literature on multilingual spoken conversation (e.g., Myers-Scotton 2006), the IM context is one in which speakers are likely to defy both constraints and norms in the interest of ease of production and to maintain a certain breeziness that is characteristic of the format. As discussed in relation to Russian and Estonian by Verschik (2007, 2016) and Zabrodskaia & Verschik (2014), morphological and phonological integration need not be interdependent; each depends to a large part on individual preference and cannot be predicted solely based on structural compatibility, particularly in computer-mediated communication, in which spontaneity and great variability are prevalent.

In the following section, we discuss relevant background context, describing characteristics of IM, Estonian-English contact in light of youth language and young people’s internet usage, and morphosyntactic and orthographical aspects of the Estonian language crucial to the analysis. In Section 3, we describe the data and method used for this study. In Section 4, we present the results of the analysis, describing in more detail how English elements are integrated orthographically and morphologically in Estonian chat conversations. We conclude with a discussion and conclusive remarks.

2 Background

2.1 Language used in instant messaging

Much of today’s communication occurs in online spaces (e.g., social media and instant messaging applications; see Thurlow & Mroczek 2011), shaped alongside rapid technological developments in the past twenty to thirty years. Both the availability of these online written formats and their technical constraints have affected language use to the point where a new language register or variant has

3 *The integration of English in Estonian teenagers' instant messages*

been identified across various languages (Collot & Belmore 1996, Androutsopoulos 2014, McCulloch 2019). Proposed labels for this new type of language include netspeak, chat language, internet slang, web-based language, internet language, or, perhaps most commonly, COMPUTER-MEDIATED COMMUNICATION (CMC; see Herring 1996, Crystal 2001, Androutsopoulos 2006).

While CMC describes the language used in technologically mediated discourse (Androutsopoulos 2013, Vettorel 2014), it is a broad umbrella term, referring to various web-based communicative practices, including not only spontaneous texted communication but also edited contexts like blogs, vlogs, and forums more generally. In this paper, we focus on a synchronous language variant used in communication between two or more speakers in rapid messaging applications (particularly Facebook Messenger and Discord), and we refer to it as INSTANT MESSAGING (IM) LANGUAGE (see Herring 2007, Tagliamonte & Denis 2008, Baron 2010).

IM language combines numerous features from written and spoken language while adapting to the technological capabilities of online spaces (e.g., Palfreyman & al Khalil 2003) and the orthographic affordances of the written format. One frequent feature involves using abbreviations and short forms (cf. Thurlow & Brown 2003, Tagliamonte & Denis 2008), such as *wdym* for 'what do you mean' and *k* for 'okay'. The technical limitations of earlier mobile device models directly influenced this characteristic: writing a text message or SMS (Hård af Segerstad 2005, Ling 2005) was time-consuming, limited in length, and costly; consequently, users found innovative ways to write messages more concisely (Crystal 2009). Although current texting applications no longer have these character limitations and allow more character variety, message shortening is still frequent. This can be partially attributed to the speed of typing a message and partially to the distinctive language register, co-creating and enacting a shared knowledge of the abbreviations' meanings (Tagliamonte 2016a).

Signs of creativity and experimentation are especially evident in IM typography (Varnhagen et al. 2010, Herring 2022). The use of upper- and lowercase letters functions similarly to prosody in spoken language: they allow the user to emphasize a message (e.g., with UPPERCASE letters often interpreted as shouting, or having an intensified meaning) or give it a specific undertone (with lowercase writings used for aesthetics or aLtErNaTiNg CaPs for expressing playfulness). Along with text capitalization, IM users employ repeated letters to echo (but not necessarily mimic) lengthened pronunciation or emphasis (e.g., *pleeeeee*, *omggggg* 'oh my god', *nopeee* 'nope'), as well as punctuation marks (e.g., repeated exclamation marks to express strong emotion) and what are known as keyboard smashes (e.g., *asdfghjkl*; see McCulloch 2019: 6) for emphatic or emotional effect.

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

While gestures and facial expressions contribute to meaning-making in spoken interaction, various visual components in IM language function in a similar way. Gawne & McCulloch (2019) describe emoticons, emojis, and animated GIFs as *digital gestures*. Depending on the context, these can accompany a written text or hold meaning independently.

As literary practices have adapted to online spaces, so have the languages – and specifically orthographies – in which IM (and other web-based) texts are produced. One example of this adaptation is the process of TRANS-SCRIPTING (Androutsopoulos 2020), where a language is respelled with alternative characters, reflecting a type of orthographic integration. Trans-scripting is often employed with languages that make use of a non-Latin script (e.g., Su 2003, Androutsopoulos 2012); however, we can observe this kind of adaptation in languages where the spelling reflects pronunciation, e.g., the word ‘nice’ written as *naiss* in Estonian.

Another distinguishing feature of IM involves the representation of colloquial, spoken language in the spelling conventions, such as *dunno* for ‘I don’t know’ and *prolly* for ‘probably’ in English. In Estonian, spelling more directly reflects the phonetic system (see Section 2.3.) and can easily be deployed to indicate the phonetic modifications arising in rapid speech. Specific, pronunciation-based variants have come into everyday use as IM usage has spread, including frequent collocations like *see on* ‘this is’ > *sen*, *ma olen* ‘I am’ > *malen*, *ma ei tea* ‘I don’t know’ > *mtea*, *mdea*. Other commonly used pronunciation spellings include *põmst* (< *põhimõtteliselt* ‘basically’), *tegelt* (< *tegelikult* ‘actually’), and *aint* (< *ainult* ‘only’).

In interactions among multilingual interlocutors, CODE-SWITCHING is frequent. Known as the alternation of two languages (or language variants) within a single discourse (Poplack 1980), code-switching can occur both intra- and intersententially. As noted earlier, the polylingualism characteristic of IM language puts the very notion of code-switching into question, as it is not always clear whether the speaker is “switching” between languages or making use of an emergent language blend. Most of the research on code-switching examines data collected from spoken language, while research on multilingual language use in the context of computer-mediated communication is in its early stages. Deuchar (2020) notes that even in spoken discourse, differentiating between a code-switched or borrowed element is less than straightforward (we explain how we tackled this in Section 3.2.). The inherent variability of code-switching (or polylinguaging) in IM practices further complicates the question. Barasa (2016), investigating code-switching in Kenyan University students’ online messages, concludes that CMC code-switching should be treated as distinct from spoken code-switching,

3 *The integration of English in Estonian teenagers' instant messages*

as certain discourse functions were identified as unique to CMC. Others investigating the pragmatic functions of code-switching have arrived at the same conclusion (Hinrichs 2006, Herring et al. 2013). Androutsopoulos (2015) further expands this theorizing by proposing the term NETWORKED MULTILINGUALISM to describe the multilingual practices in and characteristic to social network environments. A major focus of studies concerning CMC-centered multilingualism is the use of English (and its adaptation) in the context of other languages. De Decker & Vandekerckhove (2012) investigated the use of English in the MSN conversations of Flemish adolescents; Isenmann (2022) observed the use of English in Icelandic Facebook; Verheijen & van Hout (2022) examined social media messages of Dutch youth. Each of these studies report that both the use and integration of English lexical items is highly variable, with the elements (particularly phrases and longer units) sometimes remaining unadapted, while in other cases English is creatively integrated within the graphemic and morphological frames of the matrix language. The localization of English shows that CMC-centered multilingual practices should be fruitfully viewed as a form of translanguaging.

2.2 Estonian-English contact, youth language and the internet

The last three decades have crucially brought about two types of language contact in Estonia (Praakli & Koreinik 2020: 93): first, the number of languages in contact with Estonian has doubled since the country regained independence in 1991; secondly, contact between Estonian and English speakers has greatly increased. English is used as a common language with people of varying backgrounds and in online communication. English-Estonian contact is more multifaceted than it was in the past when it was mostly characterized by cultural loanwords and translations (Jõgi 2014). English functions as the international lingua franca in politics, business, and entertainment. In some situations, English is used as a vehicular language, while in others, it is drawn on as an additional communicative resource to serve various discursive purposes (Kask 2021, Praakli et al. 2022, Vihman et al. 2022).

The latest census (Statistics Estonia 2022b) reported that 48% of the population spoke English as a foreign language, making it the most widely spoken foreign language in Estonia. In the period since previous censuses (2000, 2011), the presence of English in Estonia has grown steadily. However, English proficiency varies considerably by age; it is highest among adolescents and young adults, with 85% of 15 to 29-year-olds claiming to speak English as a foreign language (Statistics Estonia 2022a). English is now taught as the first compulsory foreign language at school. The EF English Proficiency Index Education First (EF)

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

(2022) highlights Estonians' high English language proficiency, especially among young people.

The internet and other rapidly developing technology of recent decades have opened up unlimited access to English-language networks and media content. Most children in Estonia are exposed to English via media from an early age. Kalmus et al. (2022: 1027) remark that the past decade has seen significant changes in how the internet and digital devices influence the lives of children and youth, such as a substantial increase in the proportion of smartphone-using children and the amount of time they spend online each day, watching videos on YouTube or social networking on global platforms such as Instagram, TikTok, and WhatsApp. Estonian children's internet use has increased markedly, with 97% of children aged 9-17 accessing the internet every day in 2018, compared to 82% in 2010 (Sukk & Soo 2019; similar statistics are reported by Kalmus et al. 2022: 1033). The media practices of young Estonians are increasingly multilingual, with English holding a key position (Praakli et al. 2022). The use of English alongside Estonian reflects their changing linguistic habits, contexts of use, and sources of linguistic input. Although Estonian teenagers acquire English at school, the term FOREIGN LANGUAGE is not adequate for defining the position of English in their lives. Beyond classes at school, English is acquired through the internet and entertainment. The effects of English as a global language are noticeable in countries like Estonia, where (a) the relatively small population means that the production of entertainment in the national language is much more limited than in English, particularly regarding entertainment targeted at teenagers, and (b) liberal social attitudes mean that young people are given a great deal of autonomy in organizing their extracurricular time.

Hence, for today's Estonian teenagers, English is neither the L1 (first language) nor a foreign language (learned outside the environment in which it is spoken). It would be appropriate to describe it as a SECOND LANGUAGE (L2) learned partly through IMMERSION, despite the virtual nature of that immersion. Teenagers are 'digital natives' who converse even with their local friends mostly through instant messaging, social media, and smart devices (see Paus-Hasebrink et al. 2019). Although the English-language input in the immersion context is mostly virtual, for this generation, that is a seamless extension of their ordinary language use and a fairly natural context of acquisition, as well as an organic enhancement of their interactions with their peers: although a group of friends may be all L1 Estonian speakers, they will draw on English when communicating with each other because of their shared polylingual online experiences. We note that today's teenagers often feel at home in their L2 English, switching easily between Estonian and English.

3 *The integration of English in Estonian teenagers' instant messages*

In summary: the multilingualism of this generation should not be understood as “full competence in different languages” (Blommaert et al. 2005: 199), but as a core feature of their language use and communicative needs (see Section 3 for participants' proficiency). In young people's spontaneous spoken conversations (Vihman et al. 2022), computer-mediated interactions (Igav 2013, Kilp 2021), and performative videos (Praakli et al. 2022), English functions as an additional language resource, used for stylistic and pragmatic purposes. Nevertheless, data on the role of English in unedited, spontaneous, digitally mediated discourse among young people in Estonia is sparse.

2.3 Some features of Estonian morphosyntax and orthography

Estonian is a Finno-Ugric language affected by contact with Indo-European, especially German and Russian, but genetically unrelated to English (Metslang 2009). Whereas English is analytic and employs rigid word order to indicate grammatical structure, Estonian is fusional-agglutinative and makes rich use of inflectional morphology for marking grammatical relations and argument structure.

Estonian nouns take fourteen cases in singular and plural, while in English, only pronouns show any case-marking. Plurals are marked in both languages. English plurals are generally marked with an -s, though irregular plurals have various patterns. Estonian nominative plurals are marked with a regular -d affix, which is attached to the genitive stem rather than the nominative (see Table 1, final column). Importantly, the genitive always ends in a vowel, known as a THEME VOWEL (glossed as TV), but which particular vowel (*a*, *e*, *i*, or *u*) is lexically specified (Table 1 gives examples of genitive formation). Nouns ending in a vowel in the nominative case (e.g., *pere* in Table 1) often use the same form in genitive singular. Importantly, the genitive form acts as a stem for the nominative plural and most of the singular cases (e.g., allative in Table 1), and the vowel is crucial for the phonotactics of adding the case suffixes.

Partitive case, which is used to mark direct objects as well as partitivity, is formed in two ways in the singular across different declension classes, with either vowel-final forms (*õuna*, *kiike*, *tooli* in Table 1) or forms ending in -t/d (*raamatut*, *peret*). Genitives and vowel partitives may involve stem changes.

Estonian verbal morphology is agglutinative and, as in noun morphology, the stem ends in a vowel. Verb affixes mark tense, person, number, mood, and voice in their finite form or receive various infinitival markers.

The wealth of inflectional morphology used in Estonian requires speakers to make frequent choices with borrowings and code-switched words, to (a) mark

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

Table 1: Partial inflectional paradigms for Estonian nouns: five example lexemes, five paradigm cells (Nominative, Genitive, Partitive and Allative singular; Nominative plural)

	NOM. SG	GEN. SG	PAR. SG	ALL. SG	NOM. PL
apple	<i>õun</i>	<i>õuna</i>	<i>`õuna</i>	<i>õuna-le</i>	<i>õuna-d</i>
swing	<i>kiik</i>	<i>kiige</i>	<i>kiike</i>	<i>kiige-le</i>	<i>kiige-d</i>
chair	<i>tool</i>	<i>tooli</i>	<i>`tooli</i>	<i>tooli-le</i>	<i>tooli-d</i>
book	<i>raamat</i>	<i>raamatu</i>	<i>raamatu-t</i>	<i>raamatu-le</i>	<i>raamatu-d</i>
family	<i>pere</i>	<i>pere</i>	<i>pere-t</i>	<i>pere-le</i>	<i>pere-d</i>

them with morphology from the source language, (b) mark them with morphology from the matrix language, or (c) leave them unmarked. Zabrodskaia (2009) describes the gradience of morphological integration for Estonian words used in Russian (see also discussion in Vihman 2016, Kask 2019), including the question of whether or not morphology is required but also various intermediate levels between no integration and full integration. In Estonian, morphological marking involves further choices, notably how to accommodate stem changes on elements from other languages and which theme vowel to use with grammatical affixes. Because of the mixture of stem-changing and affixal morphology, integration into Estonian allows for differing degrees of integration, as described for Welsh by Deuchar (2020).

For neologisms, loanwords, or switches integrated into Estonian morphology, the default theme vowel is *-i*, used for both nouns and verbs. English orthography often provides another alternative: although the unpronounced *-e* found at the end of many words (e.g., *nice*, *smile*) does not affect phonological integration in the spoken language, it can influence how speakers accommodate inserted stems in informal written code-switching.

The two languages also differ in their orthography. Unlike English, which uses an opaque writing system, the orthography of Estonian is essentially phonetic. Phonemes are transparently mapped to letters, with a nearly one-to-one correspondence. The Estonian alphabet includes four vowels with diacritics, *Õ* pronounced /ɤ/, *Ä* /æ/, *Ö* /ø/ and *Ü* /y/. The letters C, Q, W, X, Y, and Z are only used in loanwords, foreign names or playful spellings. Y has often been used informally instead of *Ü* to denote /y/, as it is used in the closely related Finnish language. The Y used in English words can be ESTONIANIZED with various representations: J (e.g., *jess* for ‘yes’), AI (*oomaigad* ‘oh my god’), or I (*sori* ‘sorry’). In

3 *The integration of English in Estonian teenagers’ instant messages*

addition to the Z, two further sibilant phonemes are written in loanwords with letters with diacritics, Š and Ž (e.g., *šokolaad* ‘chocolate’, *džungel* ‘jungle’).

Geminates and long vowels are written with double letters, and the length distinction is morphosyntactically contrastive. On the whole, the spelling is taken to reflect the pronunciation.

3 Data and method

3.1 Participants and data

Data for the current study were extracted from a chat corpus compiled in 2020-2021 within the Teen Speak in Estonia project (Koreinik et al. 2023, Vihman et al. 2023). Institutional ethics approval was obtained from the Research Ethics Committee of the University of Tartu (nr. 319/M-13, 15.06.2020). The corpus contains 109 IM conversations between 102 unique participants aged 10-18. Over 77% of the participants self-reported that on a scale from 1 (no command at all) to 5 (fluent, with intervening points unlabelled), their English proficiency is either 4 (33.3%) or 5 (44.1%, see Table 2). The median across all age groups (10-12, 13-15, and 16-18 years of age) was 4.

Table 2: Self-reported proficiency in English

Rating	N	%
1	0	0
2	3	2.94
3	20	19.6
4	34	33.3
5	45	44.1

The chats, chosen and donated to the research project by the participants themselves, are of varying lengths, ranging from 5 to 2946 tokens ($M=547$, $SD=601$, $Mdn=294$), including links, images, emoticons, and emojis.² The total token count of the chat corpus at the time of analysis is 59,677 (including 1613 images or emojis and 110 links).

²Since the number of tokens is calculated automatically based on elements which are separated by turn switches and spaces within the turns, it is highly influenced by the use of features characteristic of IM in each chat, such as abbreviations, acronyms, emoticons and emojis, informal punctuation, shortening, etc.

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

The texts were manually annotated for language choice by a group of researchers and students of linguistics and checked for consistency by the authors. In order to decide whether a “borderline” word is a code-switch or an established loan, we held discussions, making reference to its inclusion (and labelling as colloquial) in the comprehensive online dictionary,³ its use in other corpora and the coders’ language experience. Vihman et al. (2022) report that 7.5% of the words used in the same chat corpus were English (compared to 3.3% in spoken conversations), but that this fluctuates by speaker and conversation. While this percentage is expected to be lower in longer texts, a great deal of variation exists even in shorter conversations; Figure 1 shows the length in tokens (including words, emojis and other visual items) for each participant on the x-axis, and the proportion of words in languages other than Estonian on the y-axis.

For the purpose of this study, all tokens marked as non-Estonian, along with their surrounding context, were extracted from the data. Consecutive non-Estonian words were automatically grouped into multiword expressions (compounds, phrases, and sentences). This resulted in an initial dataset with 2269 observations. After manual inspection and correction of the automatic annotations, 2260 observations remained, of which 2234 (98.85%) were in English, followed by 16 in Russian (0.71%), 8 in French (0.35%), and 2 in Finnish (0.09%). In the final dataset, only the English observations were used.

Out of the 102 participants in the chat corpus, only 3 use Estonian exclusively in their IM conversations. This means that our final dataset contains data from 99 participants, 70% of whom are female (Figure 2).

In all age groups, but especially among the youngest participants, the average rate of English usage is higher for boys (Figure 3). However, according to Wilcoxon rank-sum tests, the differences between genders are not statistically significant (all *p*-values exceed .05), although in the youngest group, this results from the presence of outliers and a small sample size.

3.1.1

3.2 Principles for coding code-switching

All observations in the dataset were coded by the authors, in frequent discussion for purposes of reliability and consistency, with regard to their functional category in the turn as well as their orthographic and morphological integration into Estonian. This allowed us to examine which English elements are

³<https://sonaveeb.ee/> (16 October, 2023). The dictionary has adopted an inclusive approach, including words such as *tšillima*, *tsillima* (from the English ‘to chill’), marked as colloquial. The orthographically unintegrated *chillima*, however, is not included.

3 The integration of English in Estonian teenagers' instant messages

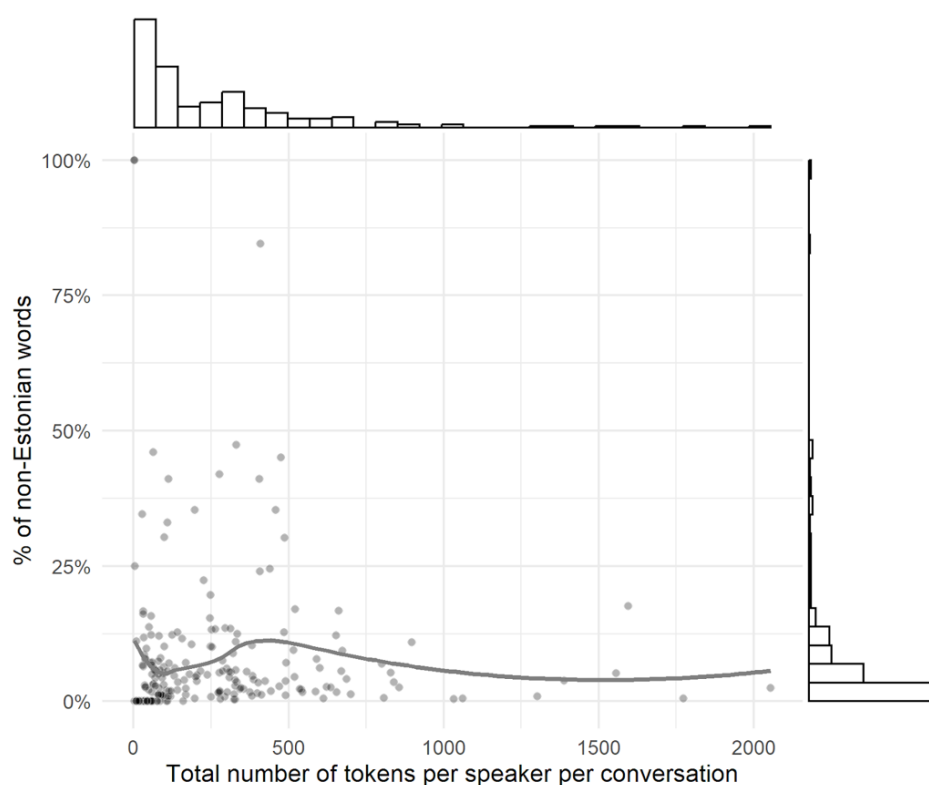


Figure 1: Percentage of non-Estonian tokens, according to the number of total tokens per speaker per conversation in the Estonian Teen Speech IM chat corpus. The histograms on the sides show the overall distribution of the number of tokens (top) and the percentage of non-Estonian words (right) in the conversations per speaker

most often used, and which linguistic expressions are integrated into Estonian. See Table 3 for a summary of the coding system used, along with examples of each category (the dataset and coding principles are also publicly available at <https://osf.io/yge27/>).

The coding of a functional category generally coincides with part-of-speech tagging (e.g., conjunctions, interjections, adverbs, nouns, verbs, adjectives), but we also categorized larger units such as fixed phrases and sentences. Although coded separately, in this study, noun phrases, adjective phrases, and verb phrases are subsumed under the corresponding general part-of-speech tags (N, A, and V, respectively) due to their infrequent use. Direct quotes (e.g., from song lyrics) are assigned a separate category.

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

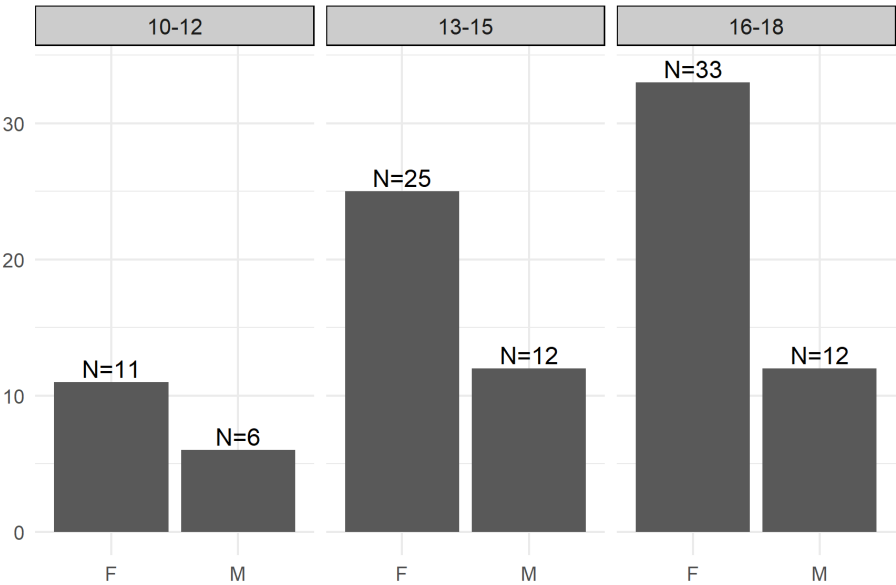


Figure 2: Unique participants (N=99) in the final dataset by gender and age group

We considered the inserted elements to be ‘orthographically integrated’ when their spelling deviates from their expected (standard or nonstandard, e.g., *bruh* as used in English-context IM language) spelling in the source language, which in this case is English. This happens when the speakers integrate English into the phonetically transparent Estonian orthography (*is veri guud* < ‘is very good’, *dunt võrri* < ‘don’t worry’, *verinaiss* < ‘very nice’). In some cases, especially with compounds, phrases, and sentences, but also with some simplex words, only part of the expression is integrated (*thankjuu* < ‘thank you’, *whateefak* < ‘what the fuck’), in which case the items were coded as exhibiting mixed orthographic integration. Abbreviations, such as *wym* ‘what [do] you mean’ and *omg* ‘oh my god’, were considered difficult to assess for (non-)integration, and they were marked as NA (not applicable).

Non-Estonian words and expressions in the chats were coded for being ‘morphologically integrated’ when they incorporated elements from Estonian inflectional (and, occasionally, derivational) morphology. This mainly concerns nouns (*bag-i-d* bag-TV-PL ‘bags’), adjectives (*clean-i-m* clean-TV-CMP ‘cleaner’), and the corresponding phrases, which allow for inflectional marking for case, number, and degree, as well as verbs (*quitt-i-sin* quit-TV-1SG.PST ‘(I) quit’) and verb phrases

3 *The integration of English in Estonian teenagers’ instant messages*

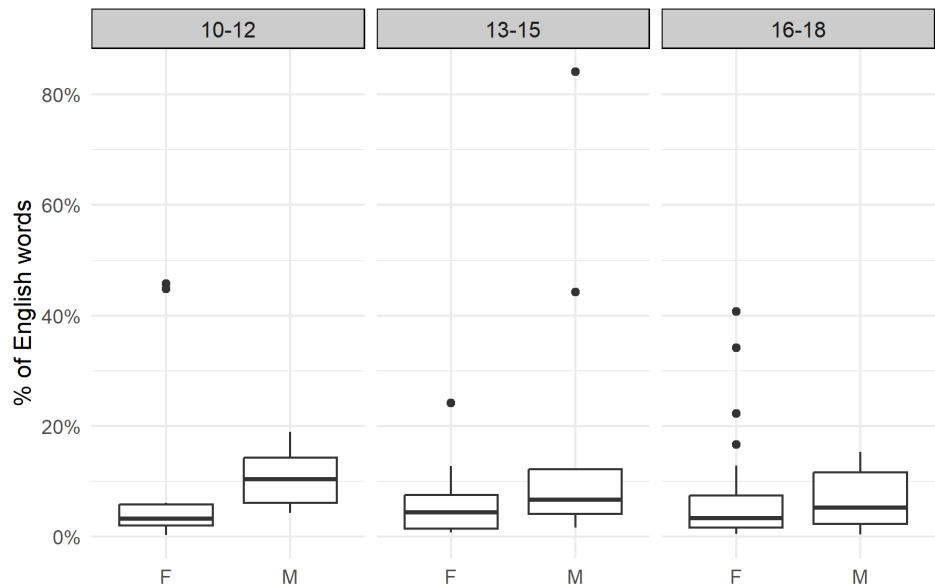


Figure 3: Percentage of English words used by the participants by gender and age group

(*reality-t face-da reality-SG.PAR face-INF* ‘to face reality’). The items were coded with NA when the grammatical function did not require any use of inflectional morphology in Estonian (e.g., the unmarked nominative singular case, imperative mood for some inflectional verb types, other parts-of-speech, such as adverbs, interjections, conjunctions, and fixed phrases and sentences). English elements were coded as not being morphologically integrated when their grammatical function would call for such marking in standard Estonian, but the speakers had either left them unmarked or used the morphology from the source language. Therefore, the binary opposition between morphologically marked and unmarked elements in our data only reflects whether the potential for morphological integration in specific contexts was realized or not.

4 Results

4.1 English code-switches by category

In this study, since we are interested in the integration of words and phrases both orthographically and morphologically into Estonian-language contexts, we are

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

Table 3: Summary of coding categories with examples. If an example contains an Estonian word with nonstandard spelling, the standard orthographic form will be presented in square brackets [:]; if the example contains an Estonian word with nonstandard morphology, the standard form will be presented in round brackets (:).

Functional category	
adjective	<i>Pizzatastic, coolcoolcool</i>
adverb	<i>prolly</i> ‘probably’, <i>btw</i> ‘by the way’
conjunction	<i>cuz</i> ‘because’, <i>tho</i> ‘though’
noun	<i>playlist, DUMBASS</i>
verb	<i>join-i-da</i> ‘to join’, <i>SPOIL-I-MA</i> ‘to spoil’
interjection	<i>brooo, Jaaa</i> ‘yeah’
question word	<i>howwwww, wut</i> ‘what’
phrase	<i>I guess, oh well</i>
quote	<i>Thank you next, I got tha moves like jagger</i>
sentence	<i>and where u at, less goo</i> ‘let’s go’
Orthographic integration	
yes	<i>oomaigad</i> ‘oh my god’, <i>kamooon</i> ‘come on’, <i>oukei</i> ‘okay’
no	<i>Alrightt, yeeee, GREEN SCREEN-I-GA</i> ‘with a green screen’
mixed	<i>scämm</i> ‘scam’, <i>cropp top-i-d</i> ‘crop tops’, <i>hope iz uki</i> ‘hope (it) is okay’
NA	<i>Ikr</i> ‘I know right’, <i>WDYM</i> ‘what do you mean’, <i>Lvl up</i> ‘level up’
Morphological integration	
yes	<i>ma tulen gym-i-st</i> ‘I’m coming from the gym’, <i>Nii hea edit-i-ja</i> ‘such a good editor’, <i>mulle ei meeldi reality-t face-da</i> ‘I don’t like to face reality’
no	<i>tahaks teha seda challenge</i> (: <i>challenge</i> ’i) värki ‘I’d want to do that challenge thing’, <i>hessi friikad on soggy</i> (: <i>soggy</i> ’d) ‘Hesburger’s fries are soggy’, <i>mingid fancy</i> (: <i>fancy</i> ’d) <i>veinipokaalid</i> ‘some fancy wine glasses’
NA	<i>i need to go niitma</i> ‘I need to go mow the lawn’, <i>ma probably votan need</i> ‘I’ll probably take these’, <i>okei see on ylinaiiss</i> ‘okay that’s really nice’

3 The integration of English in Estonian teenagers' instant messages

especially interested in intrasentential usage. Hence, we first report the proportions with which English-language segments of all kinds (including full turns and whole sentences) are used in the data and then limit our focus to sub-utterance level switched elements.⁴ We also exclude titles and quotes (especially song lyrics) from further analysis, to the extent that we were able to identify them.

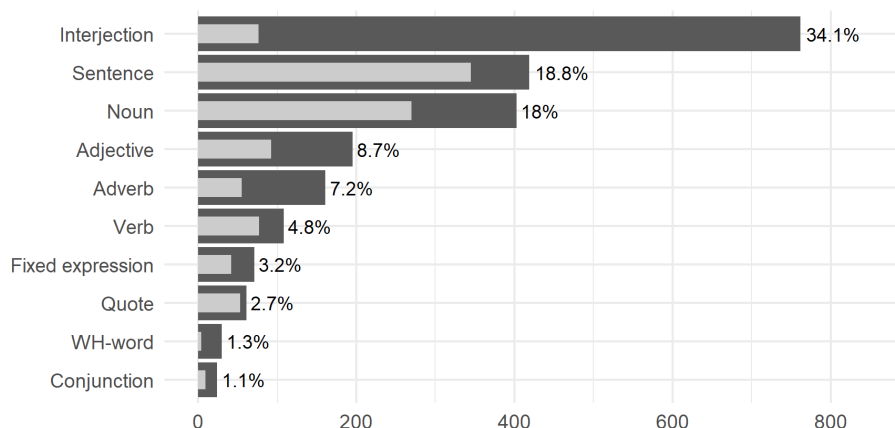


Figure 4: The categorical distribution of English expressions by tokens (dark grey bars) and standardized types (light grey bars)

As can be seen in Figure 4, 53% (N=1181) of the English expressions in our dataset occur in intersentential positions, where the segments are not syntactically linked to the preceding or succeeding context. These exclude quotes (which have both inter- and intrasentential uses) and include simplex and multiword interjections (e.g., ‘wow’, ‘oh my god’) and sentences. However, the numbers of unique standardized types (e.g., ‘please’) corresponding to the various surface forms of the interjections (e.g., *please*, *pliiis*, *pliizzzzz*) show that the participants make repeated use of a relatively small set of expressions in the interjection category. This also holds for adjectives and adverbs, where unique types make up less than half of all attested tokens. The most common types and their realizations are discussed in the following subsections.

4.2 Orthographic integration

Overall, non-integration into Estonian orthography is more common than Estonianized spelling. Nearly one-fifth (17.2%, N=374) of English elements were

⁴All linguistic examples are presented in their original typographical form.

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

coded as NA for orthographic integration, meaning that integration was not an option; these mostly include abbreviations (e.g., *btw* ‘by the way’, *lmao* ‘laughing my ass off’, *lol* ‘laugh out loud’, *wtf* ‘what the fuck’, *fr* ‘for real’). Out of the codable tokens, 83% (N=1492) were unintegrated orthographically, 13.7% (N=247) were integrated, and 3.3% (N=60) were coded as mixed. Mixed integration indicates that elements of English spelling and Estonian-based spelling were used. English spelling includes letters used only in loanwords in Estonian (e.g., Q in ‘quit’, C in ‘creepy’, Y in ‘yo’, ‘actually’), as well as the many orthographic representations in English which do not map phonemes to letters. Estonianized, or integrated, spelling in mixed words is seen, among other features discussed below, in spelled-out diphthongs (e.g., *noupp* < ‘nope’), the use of Estonian diacritics (*scämm* < ‘scam’, or *dunt vörri* < ‘don’t worry’), and Estonian vowels used in place of English spelling (*in realati* < ‘in reality’). When a speaker corrected themselves, we did not count the nonstandard spelling. Otherwise, we cannot rule out the possibility that some nonstandard spelling arises from misspelled words or unintentional typos, but the extent and systematic nature of the Estonianized spellings indicates that this is a commonly used device.

Not all speakers make equal use of the option to integrate the English elements into Estonian orthography. Among the 86 speakers (out of 99) who contributed at least 4 codable (i.e. non-NA) observations to the dataset, the integration rates ranged from 0 to 100 ($M=25.2\%$, $SD=25.5\%$, $Mdn=17\%$). Interestingly, the youngest age group (ages 10-12) made considerably less use of orthographic integration than the older age groups ($M=13.5\%$, $SD=12.4\%$, $Mdn=8.3\%$). However, the integration rates are also influenced by the functions in which the English elements are used.

As Figure 5 shows, interjections are not only the most frequent English-language elements but also the most frequently integrated orthographically. In this section, we examine interjections, with their high rate of 28.3% integrated (N=216, including mixed) spellings, verbs (18.5%, N=20), adjectives (13.8%, N=27), and nouns (6.5%, N=26). If we consider only codable items (excluding NAs), the integration rates of interjections, verbs, adjectives, and nouns are 41%, 18.5%, 14%, and 7%, respectively. A table for the data in Fig. 5 is given in the Appendix (Table A1).

4.2.1 Interjections

Most of the orthographic integration occurs in frequently used interjections (41% of codable items). The most frequent English-language interjections are shown

3 The integration of English in Estonian teenagers’ instant messages

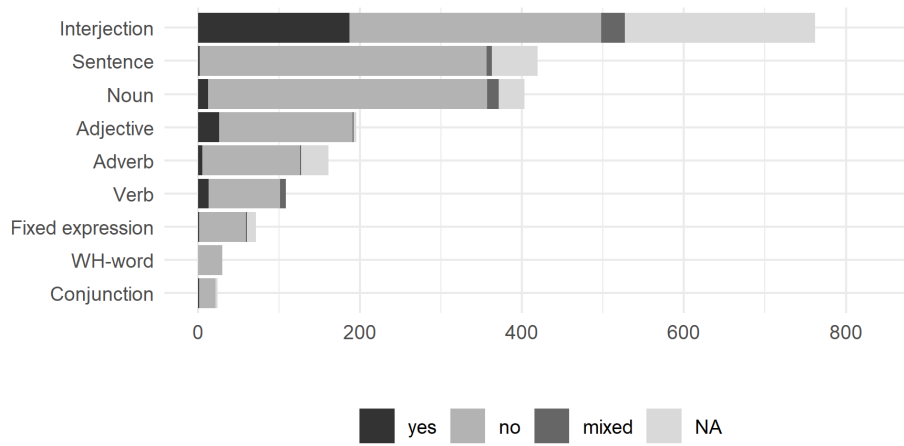


Figure 5: Orthographic integration by category (yes = orthographically integrated, no = unintegrated)

in Table 4. These include abbreviations characteristic of IM language internationally (*lol* ‘laugh[ing] out loud’, *omg* ‘oh my god’), and commonly used words in English (‘yeah’, ‘wow’, ‘sorry’). They range from three orthographic variants of ‘lol’ to 23 variants of ‘yes’, 19 variants of ‘yeah’, and 15 of ‘yup’ (the last three could also be grouped together as the affirmative particle). While most of the variation can be classified under features typical of internet language – pronunciation-based spelling, lengthening words by repeated letters or syllables, and shortening words by letter omission – we also find examples in the more frequent interjections of both integrated and non-integrated spelling variants, both within and across speakers.

Some further frequent interjections with varying forms include ‘thanks’ (20 tokens/ 9 spelling variants, e.g., *thanks*, *tanks*, *thx*, *tnx*, *tānks*, *tānx*) and ‘thank you’ (11/8, e.g., *thank you*, *tankjuu*, *thankjuuu*, *thankyouuu*, *ty*, *tānkju*), ‘oh my god’ (25/9, e.g., *oomaigad*, *omaikad*, *omg*, *omgk*, *ommmmmggggggggggggggg*), ‘bye’ (12/7, e.g., *bye*, *byeee*, *baiiii*), ‘okay’ (14/9, e.g., *okay*, *ok*, *ouk*, *oukei*, *oooook*, *kkkk*, *okeoke*), and ‘nope’ (13/5, e.g., *nope*, *noupp*, *noouupp*). Note that in expressions like ‘thanks’ and ‘oh my god’, the orthographic rendition tends to be internally consistent: the Estonian -ä- does not co-occur in our corpus with the English word-initial *th*-; similarly, the spelling *mai* for ‘my’ appears with *gad*, but never with the English spelling *god*. On the other hand, with such rampant variability, this is clearly not a hard constraint: we also find mixed orthography such as *jeah* ‘yeah’ and *thankjuuu* ‘thank you’.

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

Table 4: Orthographic integration of the most frequently used English-language interjections

Standard expression	N tokens	N spelling variants	Examples: ortho-graphically integrated, mixed	Examples: not ortho-graphically integrated or NA
‘laughing out loud’	115	3	-	<i>lol, lolol, looooooooool</i>
‘yup’	65	15	<i>jep, jap, jaaap</i>	<i>yep, yepp, yup</i>
‘yo’	53	16	<i>jo, jou, joou, joukijou, you, yoooooooo</i>	<i>yo, yoooo, yoyo</i>
‘yes’	49	23	<i>jaas, jes, jesss</i>	<i>yes, yas, yaaass, yesh, yassshshh-hhh</i>
‘yay’	39	13	<i>jee, jeii, jeje, jeiii, yeyy</i>	<i>yay, yaaaaaaaaaaaaaaaaaaaaaaaaaaaa</i>
‘please’	34	15	<i>pliiiii</i>	<i>please, plss, plsss, plz, plzzzz</i>
‘yeah’	33	19	<i>je, jee, jea, jeeaaaa, jeah</i>	<i>ye, yea, yeah, yeye</i>
‘sorry’	32	9	<i>sori, sorri</i>	<i>sorry, sorryy, sry, srryyyyy, sryyy</i>
‘aww’	28	14	<i>awii, awsii</i>	<i>aw, awh, aww</i>
‘wow’	27	5	<i>vauu, vov</i>	<i>wow, woow, wowowo</i>

3 The integration of English in Estonian teenagers' instant messages

The spelling variants of both orthographically integrated and non-integrated forms are, of course, not equally distributed and some variants are clearly more conventionalized both in terms of their overall usage frequency as well as their spread across different speakers. For example, out of the 65 uses of 'yup', 24 occur in the form of *jep* and 22 in the form of *jap*. *Jep*, in turn, is used by 15 different speakers, while *jap* is used by 10 different speakers. The other spelling variants occur 1-2 times and are also used by 1-2 speakers. However, our corpus is currently too small to make a sufficiently credible distinction between individual innovations and spellings more widely accepted in the community, especially for the expressions whose overall frequency in the corpus is lower.

4.2.2 Verbs

While only 108 tokens of English verbs and verb phrases were found in our corpus, 18.5% of these were coded as orthographically either integrated or mixed. The English verbs are frequently given a geminated stem-final consonant before Estonian inflectional morphology (e.g., *cropp-i-nd* crop-TV-APP 'cropped', *tripp-i-da* trip-TV-INF 'to trip', *lagg-a-b* lag-TV-3SG.PRS 'lags', *robb-i-b* rob-TV-3SG.PRS 'robs', *quitt-i-sin* quit-TV-1SG.PST '(I) quit', *snäpp-i-sid* snap-TV-2SG.PRS '(you) snapped'). From these same examples, we can see some lexical variation in theme vowel selection, with both the default *-i* and *-a* used.

Some other modifications are also connected to the way stems behave before inflectional endings, such as the omitted W at the end of 'follow' in example (1a) and the partially integrated 'invite' in (1b-c). Examples (1b-c), from boys of similar ages, show variation: (1b) uses a short form of the verb stem (*inva*), more adapted to Estonian morphophonology, possibly indicating loanword status, while (1c) uses the full English stem 'invite', but adopts Estonian consonant gradation, according to which the /t/ occurs in its weak form and is written with *-d*. This orthography shows adaptation to Estonian, yet the diphthong /aɪ/ is left unintegrated, with English spelling.

- (1) a. (M, 12)
 Yks [: Üks] kuulsus follo-b min-d
 one celebrity follow-3SG.PRS 1SG.PAR
 'A celebrity is following me.'
- b. (M, 13)
 Mä käsi-n sin-d inva-da
 1SG order-1SG.PRS 2SG.PAR invite-INF
 'I command [them] to invite you.'

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

- c. (M, 11)
 probleem on selle-s et sa **invid-i-d** min-d
 problem be.3SG.PRS this-SG.INE that 2SG **invite-TV-2SG.PRS** 1SG.PAR
 mängi-ma, mitte vaata-ma
 play-INF NEGwatch-INF
 ‘The problem is that you’re inviting me to play not to watch.’

4.2.3 Adjectives

Adjectives follow interjections and verbs in terms of proportion of orthographic integration (14% of codable items are integrated). The most frequently attested English adjectives are listed in Table 5, along with examples of integrated and non-integrated variants. ‘Nice’ is used much more than any other adjective, with tokens nearly evenly divided between integrated Estonian spellings (varying between one and four S’s) and unintegrated variants. ‘Nice’ (pronunciation spelling > *naiss*) is also used in compounded adjectives with Estonian intensifying modifiers meaning ‘very’ and ‘super’ (*väganaiss* ‘very nice’, *ylinaiss* ‘super-nice’) and English ones (*verinaiss* ‘very nice’). Three examples are attested in the corpus of compound adjectives formed with a prefixal form of ‘very’, two of which have integrated spelling⁵ (*verinaiss* and *verikhuul* ‘very cool’), and one retains English spelling (*verygood*); note that both examples with *veri* also have Estonianized orthography for the adjective itself, whereas *verygood* retains English orthography for both compound elements.

The adjectives ‘fine’ and ‘cool’ are attested with both integrated and unintegrated variants. Interestingly, integrated ‘cool’ (*khuul*) is spelled in such a way as to indicate orthographic *but not phonological* integration: the H represents anglicized aspiration, a phonological feature that stands out in Estonian, which has unaspirated stops.

Finally, ‘true’, ‘cringe’, ‘fucked up’, ‘hot’, ‘sad’, and ‘ugly’ are not attested with integrated variants. Even though ‘true’, ‘ugly’, and ‘cringe’ also occur in nonstandard English spelling, these deviations are not aligned with Estonian orthography.

Interestingly, the most frequently used English adjectives can often be found in similar contexts as interjections, i.e., in intersentential positions where they function as complete turns expressing immediate reactions to the interlocutor’s message. This becomes especially relevant when considering their low availability for morphological integration (see Section 4.3.).

⁵Note that the Estonianized *veri* for ‘very’ is isomorphic with *veri* ‘blood’.

3 The integration of English in Estonian teenagers’ instant messages

Table 5: Orthographic integration of the most frequently used English-language adjectives

Standard expression	N tokens	N spelling variants	Examples: integrated, mixed	Examples: not integrated or NA
‘nice’	41	12	<i>nais, naissss, naissu</i>	<i>nice, noic, niceee</i>
‘fine’	7	3	<i>fain</i>	<i>fine</i>
‘true’	7	4	-	<i>true, tru, truuuue</i>
‘cool’	6	4	<i>khuul</i>	<i>cool, coolcoolcool</i>
‘cringe’	5	2	-	<i>cring, cringe</i>
‘fucked up’	5	1	-	<i>fucked up</i>
‘hot’	5	2	-	<i>hot</i>
‘sad’	5	1	-	<i>sad</i>
‘ugly’	5	2	-	<i>ugle, uglee</i>

4.2.4 Nouns

More English nouns are used than verbs, but a smaller proportion of the nouns are orthographically integrated. Out of 371 codable (integratable) nouns, only 7% are coded as integrated or mixed. Most of these are unique items. The only orthographically integrated nouns occurring with a frequency greater than one are *spammi* ‘spam’ and *vaib* ‘vibe’. ‘Spam’ occurs once with English orthography and twice with two M’s, as shown in example (2). In both cases, the double letter m seems to be driven by the case inflection *-i*, which marks this word as either partitive (2a) or genitive (2b). The word-medial vowel is left unintegrated, but in both examples, the texter has also left diacritics off of Estonian words (*akki* pro ‘äkki’ in 2a; *parast* pro ‘pärast’ in 2b), and hence this can be interpreted as their rapid messaging style.

- (2) a. (M, 12)
 Akki [: Äkki] ta ei taht-nud
 maybe 3SG NEG want-APP spam-SG.PAR

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

spamm-i [: spämmi]

‘Maybe s/he didn’t want spam.’

b. (F, 10)

(sry **spamm-i** [: spämmi] parast [: pärast] XD)

sorry **spam-SG.GEN** because-of XD

‘Sorry because of the spam [XD= laughing emoticon].’

Nouns and verbs are examined more closely in Section 4.3.

4.3 Morphological integration

Morphological integration is not directly comparable to orthographic integration. First, morphology is not available in all contexts in the way orthographic choices usually are. Whereas over 80% of English items were marked as codable for orthography, the converse is the case for morphological integration: only 13.5% (N=294) were marked as codable. Therefore, the vast majority (86.5%, N=1879) represent elements which are unintegrated grammatically. These include interjections (described above), independent and tag phrases (e.g., *all good*, *fingers crossed*, *no pressure*, *i guess*) and sentences (e.g., *and you gave yourself a unfair advantage*, *i ate cheeseballs*, *im giving tough love*, *it do be like that*, *then we can eat butt load of pasta*), as well as words which can be seen as integrated grammatically without any additional morphology. For instance, adverbs (e.g., *literally*, *prolly* ‘probably’, *offically* ‘officially’, *kinda* ‘kind of’; see 3a) are grammatically appropriate without added morphology. Others that need no additional marking are conjunctions (3b), nouns, noun phrases and adjectives in nominative singular, which uses bare forms in Estonian and does not have form restrictions (3c), and verbs ending with a vowel when they appear in imperative singular, which uses the suffixless verb stem (3d).

(3) a. (F, 13)

Ta **officially** vihka-b min-d
3SG **officially** hate-3SG.PRS 1SG.PAR
‘S/he officially hates me.’

b. (F, 17)

ja markkus sa-i täiega vihase-ks **therefore** me pol-nud
and Markkus get-3SG.PST totally mad-SG.TR **therefore**
terve päeva rääki-nud 1PL NEG.AUX-APP whole.SG.GEN day.SG.GEN

3 *The integration of English in Estonian teenagers' instant messages*

talk-APP

'And Markkus got really angry, therefore we hadn't spoken the whole day.'

c. (F, 12)

Sa ole-d mu best friend
2SG be-3SG.PRS 1SG.GEN best friend

'You're my best friend.'

d. (M, 11)

Lihtsalt copy lingi
simply copy.2SG.IMP link.SG.GEN

'Just copy the link.'

Turning to the smaller proportion of items determined to be codable for morphological integration (N=294), the results are again complementary to those for orthographic integration: 79% (N=232) are morphologically integrated, and only 21% (N=62) lack relevant morphology.

The rates of morphological integration also vary less between different speakers than those of orthographic integration. Since there are considerably fewer codable items, there are only 24 speakers (out of 99) who contributed at least 4 codable observations to the dataset. Among them, integration rates range from 50% to 100% ($M=83.5\%$, $SD=15.3\%$, $Mdn=79.6\%$) and there seem to be no significant differences between the different age groups.

Figure 6 shows the proportions of morphologically integrated English-language elements by category. If we leave aside the code-switched elements marked as NA, we find that both nouns and verbs show equally high levels of integration: 81.4% (N=149) of integratable nouns and 81.2% (N=69) of integratable verbs are morphologically integrated (37% of nouns overall compared to 64% of verbs overall bear Estonian morphological marking, if we also consider NAs); following these are adjectives, which are more or less equally likely to be integrated (53.8%, N=14) or unintegrated. We look at each of these categories in turn. A table for this data is given in the Appendix (Table A2).

4.3.1 Nouns

Nouns are reported to be the part of speech most likely to be used in code-switching (Thomason & Kaufman 1988, Field 2002). The most probable reasons for this are the diversity of referents and the greater likelihood of semantically (and also culturally) specific words used to refer to them (Backus & Verschik 2012).

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

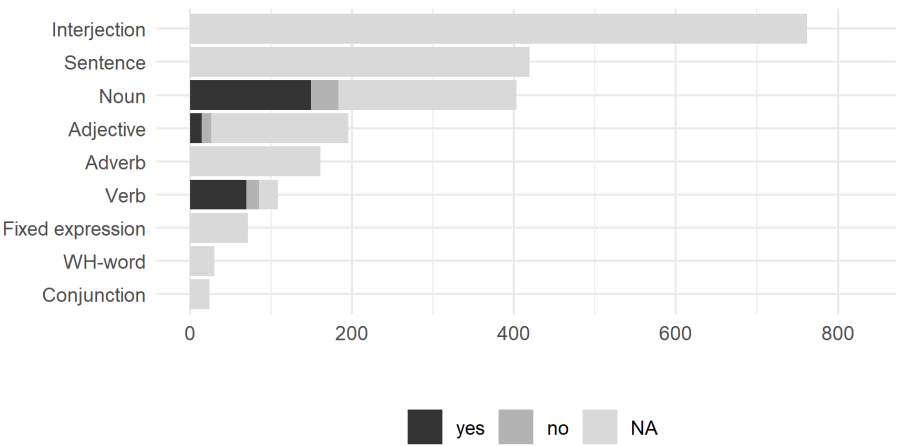


Figure 6: Morphological integration by category

In English, it has been suggested that another factor facilitating code-switching is the greater ease of grammatical integration, but in Estonian, this factor cannot play a role, as nominal morphology is complex and frequently required in Estonian clauses.

Nominative singular nouns (e.g., subjects) do not require any morphological marking, but nearly half (183) of the 403 English nouns and noun phrases in our data are coded as potentially carrying marking, i.e. are codable for morphological integration. Out of these, only 34 (18.6%) are left unmarked. Many of the unmarked nouns may be treated by the speakers like proper names when they are used in highly semantically specific ways to indicate a button or function in an app or game (*dislike*, *autotune*, *mute*), or other internet-related specific terminology (*fyp* ‘For You page’, *girl defined reaction video*, *invite to watch*, *lobby*, *Groovepad*; see example 4a). Nevertheless, proper names are also inflected in Estonian. Six nouns and one NP have English plural marking instead of integration (*flames*, *files*, *vibes*, *bikers*, *hips*, *leggings*, *religious school tingz* ‘things’). Example (4b) shows one of the few examples of an English word (denoting the name of a cafe) clearly requiring a locative ending in Estonian (illative, indicating directionality), which remains unmarked. In most such instances, the English noun is given an Estonian affix, as below in (5).

- (4) a. (M, 11)

3 The integration of English in Estonian teenagers' instant messages

tea-d join-i mu lobby lihtsalt
 know-2SG.PRS join-TV.2SG.IMP 1SG.GEN lobby simply
 'ow, just join my lobby.'

something
 missing

b. (F, 15)

Kui ma **caffeine** (: Caffeine-i) töö-le lähe-n
 When 1SG Caffeine work-SG.AL go-1SG.PRS
 'When I go and work in Caffeine.'

Hence, for the most part, English nouns, when used in an Estonian grammatical context, take some morphological marking when this is grammatically required. Plural marking takes an Estonian *-d* ending (with a theme vowel, where needed) just as often as English *-s* (*flashback-id*, *thumbnail-id*, *shader-id*, *bag-id*, *noob-id*, *cropp top-id*), including the abbreviation *bf id* 'best friends'. Plural endings in cases other than nominative can involve stem changes as well as theme vowel selection, as in (5), where the English noun, already marked for plural with the English *-s*, is double marked with the most frequent 'default' Estonian partitive plural vowel ending, *-e*.

(5) (M, 12)

Emma kiisu pane-b **cartwheel-s-e**
 Emma.SG.GEN cat put-3SG.PRS **cartwheel-PL-PL.PAR**
 'Emma's cat is doing cartwheels.'

Locative marking is shown in (6). The English word in (6a) refers to a chat message bearing the note 'seen' (used in the same way as 'read') but left unresponded to, while in (6b), the English word is very similar to the Latinate Estonian noun (*depressioon*, with two O's).

(6) a. (F, 16)

ja I would hate it kui min-d **seen-i-le** jae-takse [:
 and I would hate it if 1SG.PAR **seen-TV-ALL**
 jäetakse]
 leave-IPS
 'And I would hate it if I were left on *seen*.'

b. (M, 11)

Miks sa **depression-i-s** ole-d?
 why 2SG **depression-TV-IN** be-2SG.PRS
 'Why are you depressed?' (lit. 'in depression')

3 *The integration of English in Estonian teenagers' instant messages*

- (8) a. (M, 11)
 selle soviet **hatt-i** nimi on ushanka
 that.SG.GEN soviet **hat-TV.SG.GEN** name be.3SG.PRS ushanka
 'That Soviet hat is called *ushanka*.'
- b. (F, 16)
 A: *sul meigieemaldaja on onju*
 'You have makeup remover, right?'
 B: ei **ketchup-i-ga** vota-n [:
 NEG **ketchup-TV-SG.COM** take-1SG.PRS makeup.SG.PAR down
 võtan] meiki maha
 'No, I take makeup off with ketchup.'
- c. (F, 17)
 v [: vôi] *pop mis pane-b sin-d bad ass-i-na tund-ma*
 or pop that put-3SG.PRS you-2SG.PAR **bad ass-TV-ESS** feel-INF
 'Or pop which makes you feel like a badass.'

In our corpus, the theme vowel required to adapt consonant-final noun and verb stems to Estonian morphology is often left out, especially when the English spelling includes a silent *-e*, although in spoken language, these English words would not usually be pronounced with the phoneme /e/. As shown in examples (9a-b), considerable variation is found here (see also discussion under verbs). These examples are in successive turns of a single speaker's text, hence the variation appears to be fairly unsystematic, both across and within speakers.

- (9) a. (F, 17)
 Kolmas hea **vibe-i-ga**
 third good.SG.GEN **vibe-TV-SG.COM**
 'The third one has a good vibe.'
- b. (F, 17)
 need kõik tundu-vad hea **vibe-ga**
 they all feel-3PL.PRS good.SG.GEN **vibe-SG.COM**
 'They all seem to have a good vibe.'

4.3.2 Verbs

Verbs are much more likely than nouns to require morphological marking (see Figure 6), but *when required*, they are just as likely as nouns to actually receive

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

it in the IM corpus. For the most part, just as with the nouns, verbs are orthographically unintegrated, but morphologically integrated. Morphologically unintegrated verbs include those with English third person singular endings (*sucks, gets*) and those without morphology (*chill, stop, hand in*).

The strategies regarding the stem-final theme vowel for verbs vary. In spoken Estonian, novel verbs usually take a default *-i*. Verbs with a silent *-e* in English are optionally given a theme vowel in the corpus, as with nouns. In (10a), the theme vowel is added to English orthography (this is infrequent: the other example is *game-i-da*, game-TV-INF ‘to game’); in (10b), the theme vowel replaces the English *-e* (other examples include *leav-i-sid*, leave-TV-2SG.PST ‘(you) left’, *invit-i-sid*, invite-TV-2SG.PST ‘(you) invited’, *shar-i-da*, share-TV-INF ‘to share’). The theme vowel is not added at all in (10c; other such examples are *like-b*, like-3SG.PRS ‘likes’ and *update-me*, update-1PL.PRS ‘(we) update’). Example (10d) shows an imperative verb, which is used in Estonian as a bare stem, but always ends in a vowel: here, the English verb ‘add’ is integrated with a stem-final theme vowel despite not taking any further affixes.

- (10) a. (F, 17)
 kui kruvi-d **desolve-i-sid** või midagi
 when screw-PL **dissolve-tv-3PL.PRS** or something
 ‘When the screws dissolved or something.’
- b. (F, 12)
 mks [: miks] **delet-i-sid???**
 why **delete-TV-2SG.PRS**
 ‘Why did you delete [that]?’
- c. (M, 16)
 Aga kuule, **fade-b** ära veits
 but listen.2SG.IMP **fade-3SG.PRS** PCL slightly
 ‘But hey, (it’s) fading out a bit.’
- d. (M, 10)
 Add-i min-d
 add-TV.2SG.IMP 1SG.PAR
 ‘Add me.’

Other interesting morphological effects in verbs include the inflected use of phrasal verbs, as in the form *setup-i-tud* (past passive participle of ‘set up’), or the examples in (11a-b):

3 *The integration of English in Estonian teenagers' instant messages*

- (11) a. (M, 17)
ehmata-s lõpuks ära et ma burnout-i-sin mingi
 frighten-3SG.PST finally PCL that 1SG burn.out-TV-1SG.PST some
kahe kuu-ga
 two.SG.GEN month-SG.COM
 'It eventually scared me that I burned out in like two months.'
- b. (F, 17)
mis arva-d kas sa taha-ksi-d team up-i-da
 what think-2SG.PRS Q 2SG want-CND-2SG.PRS team.up-TV-INF
Emma kingituse tegemise-l?
 Emma.GEN present.SG.GEN making-ADE
 'Do you think you'd want to team up in getting Emma a present?'

Finally, English verb phrases and phrasal verbs may also take various forms in the data. In (12a), the speaker has derived the active past participle *nolife-nud* from the English 'no-lifer', a word used in gaming contexts to refer to a person without a social life. Example (12b) contains a mixed VP consisting of a morphologically integrated English verb with an Estonian postpositional argument (lit. 'complained on that').

- (12) a. (M, 16)
Ma ole-n lic [: lihtsalt] minecraft-i
 1SG be-1SG.PRS simply Minecraft-TV.SG.GEN no.life-APP pretty-much
nolife-nud suht
 'I have simply kind of no-lifed Minecraft.'
- b. (F, 17)
Liiga paljud inimese-d complain-i-sid selle peale
 too many person-PL complain-TV-PST.3PL this-GEN onto
 'Too many people complained about that.'

4.3.3 Adjectives

Although adjectives are inflected similarly to nouns and raise similar issues in terms of stem integration, adjectives seem to be attributed less of a central role in the clausal grammatical structure and are left unintegrated proportionally much more often than nouns. As noted in 4.2, the most frequently used adjectives are functionally similar to interjections and convey reactions, emotions, and attitudes toward the interlocutor's text. As such, they can be described as

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

engagement markers and serve an important discursive role. The data includes only 28 examples of adjectives codable for morphological integration; 16 of those are morphologically integrated.

Some examples of integrated nominative plural adjectives are *lucky-d* (lucky-PL), *hot-i-d* (hot-TV-PL). One English-derived adjective with Estonian denominal and plural affixes is *master-liku-d* (master-ful-PL). Yet plural marking is omitted on adjectives as well, as shown in (13a-c).

- (13) a. (F, 17)
 Ja ss [: siis] mingi-d **fancy** (: **fancy-d**) veinipokaali-d vb
 and then some-PL **fancy** wine.glass-PL maybe
 [: vöibolla]
 ‘And then maybe some fancy wine glasses.’
- b. (F, 17)
 hessi friika-d on **soggy** (: **soggy-d**) (...)
 Hesburger.SG.GEN fry-PL be.3PL.PRS **soggy soggy** **ass**
SOGGY ASS (: **soggy ass-id**) friika-d ‘The fries from Hesburger are
 fry-NOM.PL
 soggy [...] soggy ass fries.’
- c. (F, 17)
 burger kingi burgeri-d on kinda **questionable** thoo
 Burger King.SG.GEN burger-PL be.3PL.PRS kinda **questionable** though
 ‘The burgers from Burger King are kind of questionable though.’

Likewise, as in example (14), comparative adjectives may be marked with morphology from either language (note the nonstandard form of comparative ‘good’ in 14b):

- (14) a. (M, 17)
 pole kritiseeri-da, vöibolla ole-ks ol-nud
 NEG.AUX.CNG criticize-INF maybe AUX-CND be-APP
clean-i-m, kui rand-a ei paista-ks=ki?
clean-TV-CMP if beach-SG.PAR NEG
 be.visible-CND=CLI
 ‘There’s nothing to criticize, maybe it would have been cleaner if the beach weren’t visible?’

3 The integration of English in Estonian teenagers' instant messages

- b. (M, 14)
 se-l-i [: see oli] nagu infinity war aga **gooder**
 it-be-3SG.PRS like Infinity War but **gooder**
 'It was like Infinity War but gooder [=better].'

A few interesting examples are the prefixoid in (15a), where the adjective also takes an Estonian plural, and less frequent endings in (15b-c).

- (15) a. (M, 15)
 1 a [: aasta] töö-d ole-me puru **rich-i-d**
 1 year work-SG.PAR be-1PL.PRS totally **rich-TV-PL**
 'One year of working and we'll be filthy rich.'
- b. (M, 16)
 paljud ei tead-nud et see **limited-i-ks** lähe-b
 many NEG know-app that it **limited-TV-TR** go-3SG.PRS
 'Many [people] didn't know that it would become limited.'
- c. (F, 17)
 kurat mu-l jala-d **shave-ma-ta**
 damn 1SG-ADE leg-PL **shave-INF-AB**
 'Damn it, my legs are unshaved.'

4.4 Morphological and orthographic integration: how do they interact?

As shown above, orthographic and morphological integration seem to be relatively independent processes driven by different forces in IM chats: orthographic representation of an expression essentially provides a choice between either the source or the target language, or some nonstandard variant of either; morphological representation, however, is associated with both grammatical structure and the pragmatic need to convey meaning: it can, but does not have to follow the morphophonological rules in the written form (e.g., when choosing the output for theme vowels).

Our data includes only 21 instances where the English expressions are *both orthographically and morphologically integrated*. Such expressions are primarily derived from social media and online platforms (e.g., the noun forms *äppi* 'app.SG.GEN', *inva* 'invite.SG.GEN', *laigi* 'like.SG.GEN', *laivi-d* live-PL 'live streams', *snäppi* snap.SG.ILL 'into Snapchat', or the verb forms *tšäti-me* chat-1PL.PRS 'we

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

chat’, *follo-b* follow-3SG.PRS ‘follows’, *invidi-d* invite-2SG.PRS ‘you invite’, *draidida* trade-INF ‘to trade’). In addition, there are 17 tokens of morphologically integrated expressions which are used with mixed orthography (e.g., the noun forms *cropp topi-d* crop top-PL ‘crop tops’, *offici-s* office-SG.IN ‘in the office’, *stock foto-sid* stock photo-PL.PRT ‘stock photos’ and the verb forms *croppi-nd* crop-APP ‘cropped’, *quitti-sin* quit-1SG.PST ‘I quit’, *lagga-b* lag-3SG.PRS ‘lags’).

Where morphological integration is an option, generally, the code-switched words are left unintegrated orthographically, but where orthographic integration is found, it is only together with morphological integration (when this is an option). We also find morphological integration with abbreviations, which are not codable for orthographic integration (*dmi* ‘direct message.SG.PAR’, *bf id* ‘best friend PL’, *accu* ‘account.SG.PAR’). Both orthographic and morphological integration seem to appear on words and phrases which are more familiar, more widely used expressions, possibly on the way to becoming (or having already become) conventionalized loanwords, beyond only young people’s language and IM chats. For the most part, however, the two forms of integration operate independently and do not seem to serve the same function. Orthography is highly variable and seems to be given an optional, aesthetic or stylistic function, whereas morphological integration has more of a communicative role. The use of morphological integration tends to signal syntactic-semantic information, as well as maintaining the syntactic frame of the conversation, which tends to be maintained as Estonian.

5 Discussion and conclusion

In this study, we analyzed 2,234 observations of English elements used in a chat corpus of Estonian teens aged 10-18, aiming to identify the ways in which the participants follow or flout the grammatical, orthographic and morphological frame provided by a single language. To this aim, we clarify how English words and phrases are integrated orthographically and morphologically into (primarily) Estonian conversations. This is necessary in order to describe how this generation of digital natives and multilingual chatters frame their conversations: there is a considerable amount of English, sometimes used for functions absent elsewhere, but most multilingual messages are still given an Estonian-language grammatical structural frame. Quantitatively, the IM data include over twice as much English as the spoken data (7.5% and 3.3%, respectively [Vihman et al. 2022](#)). The teens can be said to make ample use of English, both as an implicit identity-building resource and as an important device for lexical and semantic diversity. There

3 *The integration of English in Estonian teenagers' instant messages*

is also considerable individual variation in usage, and in the use of integration, although a larger corpus is needed to reliably chart this.

Interjections make up over one-third of English units in the corpus, but their high raw token count reflects the frequent use of a relatively small set of unique expressions (e.g., 'lol', 'yo', 'wow', 'please', 'sorry'). Interjections are orthographically integrated most frequently, and their spelling is highly variable (e.g., the word 'yes' has 23 different spelling variants). Our data show that orthographically integrated expressions exhibit a combination of phonologically motivated forms, using the phonetically transparent Estonian orthography, as well as various other features common in IM, such as shortening, lengthening, and telegraphic style. While integrated orthography mostly reflects the English pronunciation (i.e. orthographically integrated, phonologically unintegrated elements), it can also indicate phonological characteristics of Estonian.

In the case of verbs, nouns, and adjectives, the form of the expression is also affected by the need to integrate the elements morphologically. This can be seen, for example, in the use of consonant gemination to express quantity alternations in inflectional forms, and in the choice of a theme vowel to bind grammatical affixes. Morphological integration overall is relatively rare in our data, since it is only relevant to these three parts of speech. However, in contexts where grammatical inflection would be expected in Estonian, over 80% of English nouns and verbs are indeed morphologically integrated. The rate of integration is considerably lower for adjectives, which are often used intersententially, providing reactions to the previous text (e.g., 'nice', 'cool', 'fine') and operating as signposts for interactional dynamics, similarly to interjections. In this function, no integration is possible. Even when it is possible, however, only 57% of adjectives are integrated, far fewer than nouns and verbs. This result is similar to Kask's (2019) analysis of English adjectives in bloggers' Estonian usage: she found that 55% of adjectives were unintegrated (see also Bahtina et al. 2021).

Two implications arise from this: first, that nouns and verbs are more central to the meaning of the clause and its argument structure than adjectives, and that teenagers chatting with their friends using IM are sensitive to this. Secondly, morphological integration varies along a scale of partial to full integration, considering the presence of stem changes, theme vowels and gemination, in addition to the addition of appropriate affixes. This underlines Deuchar's (2020) gradient view of integration.

What underlies the choice to integrate or not integrate switched elements in IM? Several factors are likely to conspire to produce these effects. English is a popular resource among Estonian youth and is used widely, in both recently conventionalized and creative ways. The IM register could be said to impose two

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

opposing mandates – to demonstrate a casual approach to language, on the one hand, flouting correct usage and punctuation, and, on the other hand, to make oneself understood, which requires a degree of specificity and conventionality. Drawing on English helps navigate this tightrope, as it increases lexical diversity while circumventing ‘correct’ language usage, which implies sticking to monolingual Estonian. Moreover, the IM register is immediate and synchronous, lending high value to speed and ease of production; this can lead to an avoidance of extraneous effort, e.g., the use of diacritics, primed language usage, or the selection of the shorter word or phrase, which may be in English or Estonian. The specific technology used also affects what is prompted and easy, with differing strategies underlying predictive texting in different applications or devices (McCulloch 2019), and the way special characters interrupt the flow of text production, but we do not have this information for the data in our study.

Without a base of comparison for the English elements used, it is difficult to say which of the words included in our data should be considered to be more conventionalized borrowings, but a potential flag for this is the presence of orthographic *and* morphological integration, as well as stability of representation across language users. The current state of the chat corpus does not allow an assessment of how much of the more common, conventionalized vocabulary remains untagged for language during our coding process. These would be words and phrases that were not included in the dataset because they were not marked as ‘English’ by the taggers, and so were excluded from the process described in Section 3.2. Borderline cases may have led to some variability in coding. For example, while several forms of ‘okay’ were manually annotated as English code-switches, many more may have been perceived as loanwords by the annotators and left untagged, as the orthographically integrated *okei* has been used in both informal speech and writing for decades. The line between code-switches and borrowings has long been noted to be difficult to pin down, as discussed by Deuchar (2020: 4; see also Bullock 2009, Poplack 2018). Phonological integration is highly variable, and morphophonological reflexes (such as mutation in Welsh and stem changes affecting consonant gradation and syllable quantity in Estonian) appear in our data in novel code-switches and older elements alike. Orthographic integration alone is not a reliable tool for measuring this aspect of phonological integration.

Despite the high proportions of English elements and the great variability in the data, the morphological integration suggests that the participants in IM conversations are positioning their conversations as rooted in Estonian, at the same time as making free use of the fluidity of the IM format. We do find great individual differences, emblematic of the high variability and changing norms in the IM register. While morphological integration suggests a prevalent single-language

3 *The integration of English in Estonian teenagers' instant messages*

frame, the data on orthographic integration casts this into doubt. Orthographic integration was found to be much more frequently available than morphological integration, but much less likely to be used with the English elements. The orthography provides an optional index which can be used to mark attitude, mastery, creativity, or conventionality, and is freely manipulated, leading to great variability. Morphological integration also shows variability (particularly because much of the English usage is novel and unconventionalized), but it plays a more fundamental role in communication and is less likely to be flouted: English nouns and verbs are preferably integrated, even while adjectives are equally likely to be integrated or not. The question of syntactic frames, grammaticality and convergence deserves closer attention: our current corpus provides several intriguing examples of syntactic mixing and agreement mismatches.

The IM data are rich with novel uses, with various research avenues to explore further. Our current sample for this study is too small to reliably assess gender- or age-group-driven differences in the preferred categories for code-switching, the integration of the expressions, or pragmatic motivation for code-switching. But differences in the dynamics of social interaction, preferences in leisure activities, language proficiency, and linguistic repertoire, between individuals as well as groups defined by age and gender, call for further investigation based on a larger corpus. A cross-linguistic comparison would also be an important direction to explore, considering the potentially similar anglophone influences in IM and spoken language among young people in many non-English speaking populations. We will be increasing the sample size of our IM corpus and hope to further explore the nuances of both formal and functional aspects of code-switching among today's digitally adept youth.

Acknowledgments

We gratefully acknowledge support from the “Estonian Language and Culture in the Digital Age (2019–2027)” R& D programme of the Estonian Ministry of Education and Research (grant nr. EKKD3, which made it possible to collect the data for Teen Speak in Estonia, and EKKD-TA16, which has supported this study), and from the Kadri, Nikolai, and Gerda Rõuk research fund at the University of Tartu. We would also like to thank the editors of this volume and two anonymous reviewers, all of whose feedback has helped us improve the paper. All remaining weaknesses, needless to say, are our own.

Finally, we extend our thanks to all the teenagers who enthusiastically participated in the project and their parents for approving and supporting their participation.

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

Abbreviations

1PL	1st person plural	ESS	essive case
3PL	3rd person plural	IMP	imperative mood
1SG	1st person singular	IN	inessive case
2SG	2nd person singular	INF	infinitive
3SG	3rd person singular	IPS	impersonal voice
AB	abessive case	NEG	negation
ADE	adessive case	PAR	partitive case
ALL	allative case	PCL	particle
APP	active past participle	PL	plural
AUX	auxiliary	PRS	present tense
CLI	clitic	PST	simple past tense
CMP	comparative	Q	question particle
CND	conditional mood	SG	singular
CNG	connegative	TV	theme vowel
COM	comitative case		

References

- Androutsopoulos, Jannis. 2005. Research on youth-language. In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier & Peter Trudgill (eds.), *Sociolinguistics / Soziolinguistik*, vol. 2 (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science 3), 1496–1505. Berlin: De Gruyter Mouton. DOI: [10.1515/9783110171488.2.8.1496](https://doi.org/10.1515/9783110171488.2.8.1496).
- Androutsopoulos, Jannis. 2006. Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics* 10(4). 419–438. DOI: [10.1111/j.1467-9841.2006.00286.x](https://doi.org/10.1111/j.1467-9841.2006.00286.x).
- Androutsopoulos, Jannis. 2012. “Greeklish”: Transliteration practice and discourse in the context of computer-mediated digraphia. In Alexandra Jaffe, Jannis Androutsopoulos, Mark Sebba & Sally Johnson (eds.), *Orthography as social action: Scripts, spelling, identity and power*, 359–392. Berlin: De Gruyter Mouton.
- Androutsopoulos, Jannis. 2013. Code-switching in computer-mediated communication. In Susan C. Herring, Dieter Stein & Tuija Virtanen (eds.), *Pragmatics of computer-mediated communication*, 667–694. Berlin: De Gruyter Mouton.
- Androutsopoulos, Jannis (ed.). 2014. *Mediatization and sociolinguistic change*. Berlin: De Gruyter.

3 *The integration of English in Estonian teenagers' instant messages*

- Androutsopoulos, Jannis. 2015. Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism* 19(2). 185–205. DOI: [10.1177/1367006913489198](https://doi.org/10.1177/1367006913489198).
- Androutsopoulos, Jannis. 2020. Trans-scripting as a multilingual practice: The case of Hellenised English. *International Journal of Multilingualism* 17(3). 286–308. DOI: [10.1080/14790718.2020.1766053](https://doi.org/10.1080/14790718.2020.1766053).
- Backus, Ad & Anna Verschik. 2012. Copyability of (bound) morphology. In Lars Johanson & Martine Robbeets (eds.), *Copies versus cognates in bound morphology*, 123–149. Leiden: Brill. DOI: [10.1163/9789004230477_007](https://doi.org/10.1163/9789004230477_007).
- Bahtina, Daria, Helin Kask & Anna Verschik. 2021. English adjectives and Estonian nouns: Looking for agreement? *Frontiers in Psychology* 12. 1–14. DOI: [10.3389/fpsyg.2021.735232](https://doi.org/10.3389/fpsyg.2021.735232).
- Barasa, Sandra. 2016. Spoken code-switching in written form? Manifestation of code-switching in computer mediated communication. *Journal of Language Contact* 9(1). 49–70. DOI: [10.1163/19552629-00901003](https://doi.org/10.1163/19552629-00901003).
- Baron, Naomi S. 2010. Discourse structures in instant messaging: The case of utterance breaks. *Language@Internet* 7. 4. <https://scholarworks.iu.edu/journals/index.php/li/article/view/37586>.
- Bessett, Ryan Matthew. 2017. Exploring the phonological integration of lone other-language nouns in the Spanish of Southern Arizona. *University of Pennsylvania Working Papers in Linguistics* 23(2). 31–39. <http://repository.upenn.edu/pwpl/vol23/iss2/5>.
- Blommaert, Jan, James Collins & Stef Slembrouck. 2005. Spaces of multilingualism. *Language & Communication*. Multilingualism and diasporic populations 25(3). 197–216. DOI: [10.1016/j.langcom.2005.05.002](https://doi.org/10.1016/j.langcom.2005.05.002).
- Bullock, Barbara E. 2009. Phonetic reflexes of code-switching. In Almeida Jacqueline Toribio & Barbara E. Bullock (eds.), *The Cambridge handbook of linguistic code-switching* (Cambridge Handbooks in Language and Linguistics), 163–181. Cambridge: Cambridge University Press.
- Cheshire, Jenny. 2005. Age- and generation-specific use of language. In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier & Peter Trudgill (eds.), *Sociolinguistics / Soziolinguistik*, vol. 1 (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science 3), 1552–1563. Berlin: De Gruyter Mouton. DOI: [10.1515/9783110858020-095](https://doi.org/10.1515/9783110858020-095).
- Collot, Milena & Nancy Belmore. 1996. A new variety of English: Electronic language. In Susan C. Herring (ed.), *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives* (Pragmatics & Beyond New Series), 13–28. Philadelphia: John Benjamins.

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

- Crystal, David. 2001. *Language and the Internet*. Cambridge: Cambridge University Press.
- Crystal, David. 2009. *Txtng: The Gr8 Db8*. Oxford: Oxford University Press.
- De Decker, Benny & Reinhild Vandekerckhove. 2012. English in Flemish adolescents' computer-mediated discourse: A corpus-based study. *English World-Wide* 33(3). 321–352. DOI: [10.1075/eww.33.3.04ded](https://doi.org/10.1075/eww.33.3.04ded).
- Deuchar, Margaret. 2020. Code-switching in linguistics: A position paper. *Languages* 5(2). 22. DOI: [10.3390/languages5020022](https://doi.org/10.3390/languages5020022).
- Eckert, Penelope. 1997. Why ethnography? In Ulla-Britt Kotsinas, Anna-Brita Stenström & Anna-Malin Karlsson (eds.), *Ungdomsspråk i Norden: Föredrag från ett forskarsymposium [Youth Language in the Nordics: Papers from a research symposium]*, 52–62. Stockholm: Stockholm University.
- Education First (EF). 2022. *English Proficiency Index (EPI)*. <https://www.ef.com/wwen/epi/>.
- Field, Fredric William. 2002. *Linguistic borrowing in bilingual contexts*. Amsterdam: John Benjamins Publishing Company.
- García, Ofelia. 2009. *Bilingual education in the 21st century: A global perspective*. Chichester: Wiley-Blackwell.
- García, Ofelia & Li Wei. 2014. *Translanguaging: Language, bilingualism and education*. London: Palgrave Macmillan.
- Gawne, Lauren & Gretchen McCulloch. 2019. Emoji as digital gestures. *Language@Internet* 17. 2. <https://scholarworks.iu.edu/journals/index.php/li/article/view/37786>.
- Hård af Segerstad, Ylva. 2005. Language in SMS – A socio-linguistic view. In Richard Harper, Leysia Palen & Alex S. Taylor (eds.), *The inside text: Social, cultural and design perspectives on SMS* (Computer Supported Cooperative Work 4), 33–51. New York: Springer. DOI: [10.1007/1-4020-3060-6_3](https://doi.org/10.1007/1-4020-3060-6_3).
- Herring, Susan C. 1996. *Computer-mediated communication* (Pragmatics & Beyond New Series 39). Amsterdam: John Benjamins.
- Herring, Susan C. 2007. A faceted classification scheme for computer-mediated discourse. *Language@Internet* 4. 1. <https://scholarworks.iu.edu/journals/index.php/li/article/view/37562>.
- Herring, Susan C. 2022. Grammar and electronic communication. In Carol A. Chapelle (ed.), *The Encyclopedia of applied linguistics*, 1–9. Hoboken: John Wiley & Sons.
- Herring, Susan C., Dieter Stein & Tuija Virtanen (eds.). 2013. *Pragmatics of computer-mediated communication*. Berlin: De Gruyter Mouton.

3 The integration of English in Estonian teenagers' instant messages

- Hinrichs, Lars. 2006. *Codeswitching on the Web: English and Jamaican Creole in E-mail Communication* (Pragmatics & Beyond New Series 147). Amsterdam: John Benjamins.
- Igav, Reet. 2013. *Inglise-eesti koodikopeerimine Facebooki vestlustes* [English-Estonian code-copying in Facebook conversations]. Tallinn: Tallinn University. (MA thesis).
- Isenmann, Vanessa Monika. 2022. *Icelandic digital practices on Facebook: Language use in informal online communication*. Reykjavík: University of Iceland. (Doctoral dissertation). <https://hdl.handle.net/20.500.11815/3666>.
- Jõgi, Aino. 2014. *Inglise päritolu sõnad eesti keeles* [Words in Estonian with English etymology]. Tallinn: Estonian Language Foundation.
- Jørgensen, Annette Myre. 2013. Emotions and vocatives in Spanish teenage talk: Emotions expressed through the vocative discourse marker *tío/a* in Madrid teenage talk. In Ana Marta González (ed.), *The emotions and cultural analysis*, 145–162. London: Routledge.
- Jørgensen, J. Normann. 2008. Polylingual languaging around and among children and adolescents. *International Journal of Multilingualism* 5(3). 161–176. DOI: [10.1080/14790710802387562](https://doi.org/10.1080/14790710802387562).
- Kalmus, Veronika, Marit Sukk & Kadri Soo. 2022. Towards more active parenting: Trends in parental mediation of children's internet use in European countries. *Children & Society* 36(5). 1026–1042. DOI: [10.1111/chso.12553](https://doi.org/10.1111/chso.12553).
- Kask, Helin. 2019. To agree or not to agree? English adjectives in Estonian-English bilingual blogs and vlogs. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 10(2). 85–123. DOI: [10.12697/jeful.2019.10.2.06](https://doi.org/10.12697/jeful.2019.10.2.06).
- Kask, Helin. 2021. *English-Estonian code-copying in Estonian blogs and vlogs*. Tallinn: Tallinn University. (Doctoral dissertation). <http://www.digar.ee/id/nlib-digar:635825>.
- Kilp, Geidi. 2021. Code-copying in Estonian-English-Japanese Facebook communication: A usage-based approach. In Anna Verschik (ed.), *Multilingual practices in the Baltic countries*, 176–211. Tallinn: Tallinn University Press.
- Koreinik, Kadri, Aive Mandel, Maarja-Liisa Pilvik, Kristiina Praakli & Virve-Anneli Vihman. 2023. Outsourcing teenage language: A participatory approach for exploring speech and text messaging. *Linguistics Vanguard* 9(4). 389–398. DOI: [10.1515/lingvan-2021-0152](https://doi.org/10.1515/lingvan-2021-0152).
- Kotsinas, Ulla-Britt. 2004. *Ungdomsspråk* [Youth language]. Uppsala: Hallgren & Fallgren.

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

- Leppänen, Sirpa. 2007. Youth language in media contexts: Insights into the functions of English in Finland. *World Englishes* 26(2). 149–169. DOI: [10.1111/j.1467-971X.2007.00499.x](https://doi.org/10.1111/j.1467-971X.2007.00499.x).
- Ling, Rich. 2005. The sociolinguistics of SMS: An analysis of SMS use by a random sample of Norwegians. In Rich Ling & Per E. Pedersen (eds.), *Mobile Communications: Re-negotiation of the Social Sphere* (Computer Supported Cooperative Work), 335–349. London: Springer.
- McCulloch, Gretchen. 2019. *Because Internet: Understanding the new rules of language*. New York: Riverhead Books.
- Metslang, Helle. 2009. Estonian grammar between Finnic and SAE: Some comparisons. *Language Typology and Universals* 62. 49–71. DOI: [10.1524/stuf.2009.0004](https://doi.org/10.1524/stuf.2009.0004).
- Myers-Scotton, Carol. 2006. *Multiple voices: An introduction to bilingualism*. Oxford: Blackwell.
- Nortier, Jacomine. 2018. Language and identity practices among multilingual Western European youths. *Language and Linguistics Compass* 12(5). e12278. DOI: [10.1111/lnc3.12278](https://doi.org/10.1111/lnc3.12278).
- Palfreyman, David & Muhamed al Khalil. 2003. “A funky language for teenzz to use.” Representing Gulf Arabic in instant messaging. *Journal of Computer-Mediated Communication* 9(1). DOI: [10.1111/j.1083-6101.2003.tb00355.x](https://doi.org/10.1111/j.1083-6101.2003.tb00355.x).
- Paus-Hasebrink, Ingrid, Jasmin Kulterer & Philip Sinner. 2019. The role of media within young people’s socialisation: A theoretical approach. In Ingrid Paus-Hasebrink, Jasmin Kulterer & Philip Sinner (eds.), *Social inequality, childhood and the media: A longitudinal study of the mediatization of socialisation* (Transforming Communications – Studies in Cross-Media Research), 45–75. Cham: Springer.
- Poplack, Shana. 1980. Sometimes I’ll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics* 18(7/8). 581–618. DOI: [10.1515/ling.1980.18.7-8.581](https://doi.org/10.1515/ling.1980.18.7-8.581).
- Poplack, Shana. 2018. *Borrowing: Loanwords in the speech community and in the grammar*. Oxford: Oxford University Press.
- Praakli, Kristiina & Kadri Koreinik. 2020. Keelemuutus vs. Sotsiolingvistiline muutus: Eesti keele sotsioperioodid re-revisited [Language change vs sociolinguistic change: Socioperiods in Estonian, revisited]. *Keel ja Kirjandus* 11. 915–934. DOI: [10.54013/kk756a1](https://doi.org/10.54013/kk756a1).
- Praakli, Kristiina, Mari-Liis Korkus, Aive Mandel, Elisabeth Kaukonen, Annika Kängsepp, Triin Aasa, Kristel Algreve, Helen Eriksoo, Marion Mägi, Getri Tomson & Liina Lindström. 2022. “Mis keeles ma räägin, I don’t know”: Eesti sisuloojate inglise keele kasutusest YouTube’is [“What language I’m speaking in,

3 The integration of English in Estonian teenagers' instant messages

- I don't know": The use of English by Estonian content creators on YouTube]. *Philologia Estonica Tallinnensis* 7. 263–290. DOI: [10.22601/PET.2022.07.10](https://doi.org/10.22601/PET.2022.07.10).
- Quist, Pia. 2010. The sociolinguistic study of youth and multilingual practices in Denmark: An overview. In Pia Quist & Bente Ailin Svendsen (eds.), *Multilingual urban Scandinavia: New linguistic practices*, 6–11. Bristol: Multilingual Matters.
- Schoonen, Rob & René Appel. 2005. Street language: A multilingual youth register in the Netherlands. *Journal of Multilingual and Multicultural Development* 26(2). 85–117. DOI: [10.1080/01434630508668399](https://doi.org/10.1080/01434630508668399).
- Schuring, Melissa & Eline Zenner. 2022. English from scratch: Preadolescents' developing use of English lexical resources in Belgian Dutch. *Frontiers in Communication* 6. 788768. DOI: [10.3389/fcomm.2021.788768](https://doi.org/10.3389/fcomm.2021.788768).
- Statistics Estonia. 2022a. *Demographic and ethno-cultural characteristics of the population*. <https://rahvaloendus.ee/en/results/demographic-and-ethno-cultural-characteristics-of-the-population>.
- Statistics Estonia. 2022b. *Population census: 76% of Estonia's population speak a foreign language*. <https://rahvaloendus.ee/en/news/population-census-76-estonias-population-speak-foreign-language>.
- Stenström, Anna-Brita, Gisle Andersen & Ingrid Kristine Hasund. 2002. *Trends in teenage talk: Corpus compilation, analysis and findings*. Amsterdam: John Benjamins.
- Su, Hsi-Yao. 2003. The multilingual and multi-orthographic Taiwan-based internet: Creative uses of writing systems on college-affiliated BBSs. *Journal of Computer-Mediated Communication* 9(1). DOI: [10.1111/j.1083-6101.2003.tb00357.x](https://doi.org/10.1111/j.1083-6101.2003.tb00357.x).
- Sukk, Marit & Kadri Soo. 2019. Summary. In Veronika Kalmus, Roosmarii Kurvits & Andra Siibak (eds.), *EU Kids Online'i Eesti 2018. aasta uuringu esialgsed tulemused [Preliminary findings of the EU Kids Online 2018 Estonian survey]*, 1–8. Tartu. DOI: [10.13140/RG.2.2.32549.55526](https://doi.org/10.13140/RG.2.2.32549.55526).
- Tagliamonte, Sali A. 2016a. So sick or so cool? The language of youth on the internet. *Language in Society* 45(1). 1–32. DOI: [10.1017/S0047404515000780](https://doi.org/10.1017/S0047404515000780).
- Tagliamonte, Sali A. 2016b. *Teen talk: The language of adolescents*. Cambridge: Cambridge University Press.
- Tagliamonte, Sali A. & Derek Denis. 2008. Linguistic ruin? LOL! Instant messaging and teen language. *American Speech* 83(1). 3–34. DOI: [10.1215/00031283-2008-001](https://doi.org/10.1215/00031283-2008-001).
- Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press. <https://www>.

Virve-Anneli Vihman, Mari-Liis Korkus, Maarja-Liisa Pilvik & Kristiina Praakli

ucpress.edu/books/language-contact-creolization-and-genetic-linguistics/paper.

- Thurlow, Crispin & Alex Brown. 2003. Generation Txt? The sociolinguistics of young people's text-messaging. *Discourse Analysis Online* 1(1). 30. <https://extra.shu.ac.uk/daol/articles/v1/n1/a3/thurlow2002003-01.html>.
- Thurlow, Crispin & Kristine Mroczek (eds.). 2011. *Digital discourse: Language in the new media*. Oxford: Oxford University Press.
- Varnhagen, Connie K., G. Peggy McFall, Nicole Pugh, Lisa Routledge, Heather Sumida-MacDonald & Trudy E. Kwong. 2010. Lol: New language and spelling in instant messaging. *Reading and Writing* 23(6). 719–733. DOI: [10.1007/s11145-009-9181-y](https://doi.org/10.1007/s11145-009-9181-y).
- Verheijen, Lieke & Roeland van Hout. 2022. Manifold code-mixing in computer-mediated communication: The use of English in Dutch youths' informal online writing. *Ampersand* 9. 100091. DOI: [10.1016/j.amper.2022.100091](https://doi.org/10.1016/j.amper.2022.100091).
- Verschik, Anna. 2007. Multiple language contact in Tallinn: Transfer B2>/A1 or B1>/A2? *International Journal of Bilingual Education and Bilingualism* 10(1). 80–103. DOI: [10.2167/beb372.0](https://doi.org/10.2167/beb372.0).
- Verschik, Anna. 2016. Mixed copying in blogs: Evidence from Estonian-Russian language contacts. *Journal of Language Contact* 9(1). 186–209. DOI: [10.1163/19552629-00901008](https://doi.org/10.1163/19552629-00901008).
- Vettorel, Paola. 2014. *English as a lingua franca in wider networking: Blogging practices*. Berlin: De Gruyter Mouton.
- Vihman, Virve-Anneli. 2016. Code-switching in emergent grammars: Verb marking in bilingual children's speech. *Philologia Estonica Tallinnensis* 1. 154–172. DOI: [10.22601/PET.2016.01.10](https://doi.org/10.22601/PET.2016.01.10).
- Vihman, Virve-Anneli, Maarja-Liisa Pilvik, Aive Mandel, Annika Kängsepp, Mari Aigro, Kadri Koreinik, Kristiina Praakli & Liina Lindström. 2023. *Estonian Teen Language Corpus v.1.0*. DOI: [10.23673/re-455](https://doi.org/10.23673/re-455).
- Vihman, Virve-Anneli, Kristiina Praakli, Maarja-Liisa Pilvik & Mari-Liis Korkus. 2022. Kas noored on inglise keelele 'obsessed'? Millest räägivad korpusandmed? [Are young people "obsessed" with English? What do corpus data tell us?] *Philologia Estonica Tallinnensis* 7. 292–321. DOI: [10.22601/PET.2022.07.11](https://doi.org/10.22601/PET.2022.07.11).
- Zabrodskaia, Anastassia. 2009. *Russian-Estonian language contacts: Grammatical aspects of language use and change*. Tallinn: Tallinn University. (Doctoral dissertation). <http://www.digar.ee/id/nlib-digar:55195>.
- Zabrodskaia, Anastassia & Anna Verschik. 2014. Morphology of Estonian items at the interface of Russian-Estonian language contact data. *Sociolinguistic Studies* 8(3). 449–474. DOI: [10.1558/sols.v8i3.25505](https://doi.org/10.1558/sols.v8i3.25505).

Chapter 4

Cued gender assignment strategies and L1 effects in the acceptance of mixed Basque-Spanish DPs

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

Research on mixed determiner phrases (DPs) containing a gendered determiner from one language and a noun from another language has highlighted the existence of individual variation in gender assignment strategies (GAS), particularly when combining languages with and without gender distinctions. Such is the case of Basque-Spanish DPs, which require the assignment of masculine or feminine gender to ungendered Basque nouns (N_B) when combined with a Spanish determiner (Det_S).

This chapter investigates the GAS that drive the acceptability of mixed DPs. Specifically, it examines whether Basque-Spanish bilinguals rely on the phonological criterion (gender according to N ending), the analogical criterion (gender of the translation equivalent), or rather on (masculine/feminine) default strategies in the acceptance of $Det_S N_B$ structures. Additionally, it explores the influence of the bilingual profile –Spanish-dominant (L1S) vs Basque-dominant (L2S)– on these preferences.

Twenty-two early bilinguals (11 L1S, 11 L2S) with high proficiency in Basque and Spanish participated in an acceptability judgment task. Based on a 5-point Likert scale they rated 40 experimental sentences with mixed DPs, controlled for analogical and phonological gender and designed according to four conditions: analogical match, phonological match, double match and no match between the gender feature of Det_S and N_B . Statistical analyses, including linear mixed models, were used to assess the effects of condition, bilingual group, and determiner gender.

Participants generally accepted mixed DPs. First, there was no strong preference for a masculine or feminine default strategy; rather, responses indicated an interplay between analogical and phonological cues. Second, across the four conditions

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

tested, structures in which the gender of the determiner met both phonological and analogical criteria received higher ratings, whereas the least accepted structures were those that met neither criterion. Third, besides the general similarity in the behaviour of the two bilingual groups, L1S bilinguals showed a more pronounced rejection of mixed DPs in the non-matching condition, while L2S bilinguals showed quite homogeneous responses across conditions. Fourth, the non-matching condition was the only one in which some participants (only the L1S group) showed a (slight) preference for feminine over masculine DPs.

In line with the variability observed in previous production and acceptability studies, the current findings confirm that gender assignment in mixed Basque-Spanish DPs is guided by cued GAS rather than default strategies. Both analogical and phonological strategies are active, with bilingual profiles, and even tasks, influencing their relative strength.

1 Introduction

Alternating between unilingual discourse in either two (or more) languages and bilingual discourse, which includes elements of more than one language, is a common praxis in many bilinguals' daily life, especially when interacting with interlocutors of similar linguistic backgrounds (Bullock 2009). This switching from one language to another in discourse has received different names such as *code switching* (henceforth CS), *code mixing*, *code alternation*, etc. The view on CS is changing in recent times and moving from being considered as an indication of (many) bilinguals' incapacity to maintain unilingual speech, to a communicative resource that only bilinguals can benefit from, and use mostly voluntarily, in spontaneous conversation. CS can be manipulated, that is, "inhibited", "controlled", "forced" or "judged", as has been shown across studies designed specifically to test bilinguals' behaviour and/or attitudes towards CS. Thus, psycholinguistic research has provided relevant findings related to the processes involved in, for example, free switching, as compared to involuntary switching produced under cued instructions in experimental settings. Although some authors consider that "these studies together underscore the importance of understanding CS speech, as well as the implications of CS speech for bilingual language representations" (Xu et al. 2021: 2), it has been found that processing natural CS and artificial CS require different brain areas, and that unlike involuntary CS, voluntary CS does not necessarily incur a switching cost (Blanco-Elorrieta et al. 2018).

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

Individual variation in CS has been widely attested, which confirms that not all bilinguals are producers of CS in a similar way, with the same frequency, or in the same communicative situations. For instance, the same speaker may code-switch more frequently when interacting with regular CS producers than with interlocutors who are not, whilst (s)he may not code-switch at all in conversations with interlocutors, monolinguals or not, who regularly communicate in unilingual speech or have negative attitudes towards CS (Blokzijl et al. 2017, Lipski 2016, Parafita Couto, Deuchar, et al. 2015).

A growing body of literature recognizes the importance of studying the linguistic restrictions involved in phrases containing elements with different (lexical, grammatical, phonological) features in the corresponding languages they come from. Under the belief that the so called “conflict sites” (Poplack & Meechan 1998) can provide relevant evidence for the study of the architecture of mixed utterances from a linguistic perspective, the mixed determiner phrase (DP) is the syntactic structure that has drawn the attention of most CS researchers, themselves coming from very different theoretical and methodological approaches and which has offered the most relevant findings on the regularities and patterns existing in both child and adult CS production (Beatty-Martínez & Dussias 2019, Cantone & Müller 2008, Chan 2008, Jorschick et al. 2011, Licerias et al. 2008). In these studies, special attention has been paid, for example, to mixed DPs involving languages that differ in word order or in a particular feature such as gender (e.g. one language without vs. another with two-three M(asculine)/F(eminine)/Neuter gender distinctions). For an updated overview on theoretical and methodological issues in CS studies see Munarriz-Ibarrola et al. (2018) and Parafita Couto et al. (forthcoming).

However, CS can be investigated at receptive level as well. Exposure to both unilingual and mixed discourse is a common experience for many individuals living in bilingual communities. Similarly to reported in production studies, individual variation has been observed in the degree of exposure to CS, as well as in the degree of acceptance (Lipski 2016, Parafita Couto, Deuchar, et al. 2015). Thus, the (passive and/or active) use of CS can be considered a component of the linguistic competence of individuals, bilingual or not, living in bilingual communities. Moreover, as a feature that may allow identification of language dominance (Heredia & Altarriba 2001), CS has become a phenomenon of interest in socio- and psycholinguistics. Therefore, it is worth studying the linguistic (e.g. typological distance between the languages involved), the psycholinguistic (participants’ profile, age of acquisition) as well as the sociolinguistic factors (community type, socio-economic status, prestige) affecting CS use. See Beatty-Martínez & Dussias (2019), Blokzijl et al. (2017) and Parafita Couto & Gullberg (2019).

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

In this regard, research on mixed determiner phrases (DP) has provided a fruitful literature on bilinguals' tendencies and strategies in the configuration of such structures, as well as on the assignment or reassignment of gender features to (un)gendered nouns inserted in DP phrases containing a gender specified determiner from the other language (Bellamy & Parafita Couto 2022). For instance, English-Spanish mixed DPs may result from the combination of a Spanish determiner (Det_S) and an English noun (N_E) such in (1a), or of an English determiner (Det_E) and a Spanish noun (N_S) (1b). The expressive and receptive use of mixed DP requires that the bilingual speaker “decides” on the language that will provide each lexical element in the mixed structure. Although theoretically both options are possible, there is robust evidence coming from early and adult spontaneous production, as well as from acceptability judgment tasks, which has demonstrated that in many different communities Det_SN_E mixed DPs (1a) are more frequent than the reverse Det_EN_S (1b) (Herring et al. 2010, Jake et al. 2002, Liceras et al. 2008, Moro Quintanilla 2014, Parafita Couto et al. forthcoming, Valdés Kroff 2016).

- (1) a. *la/el house*
 $\text{the}_{S,F}/\text{the}_{S,M}$ house E. ungendered
 ‘the house’
 b. *the casa*
 the E. ungendered house $_{S,F}$
 ‘the house’

Option (1a) requires an additional adjustment in gender features between the lexical insertion of the N_E *house* and the Det_S *la/el*, which according to the Spanish grammar agrees in gender and number features with the lexical category that it determines. Thus, in order to evaluate the acceptance of the Det_S the bilingual speaker needs to decide whether to attribute M or F gender to the (originally) ungendered N_E *house*.

The aim of this study is to gain further understanding of bilinguals' gender assignment strategies (GAS) in mixed DPs by providing new acceptability data of mixed Basque-Spanish DP structures by two groups of bilinguals, who are regular and highly functional users of these two languages: S(panish) a gendered language, and B(asque), a non-gendered language. Similarly to English-Spanish bilinguals, Basque-Spanish bilinguals need to assign some gender value to each Basque noun (N_B) inserted in mixed DPs containing a Det_S (2).

- (2) a. *la etxe*

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

the_{S,F} house_{B,ungendered} (Spanish: ‘casa’_F)
 ‘the house’

b. el etxe

the_{S,M} house_{B,ungendered} (Spanish: ‘casa’_F)
 ‘the house’

To that aim, the remaining part of the chapter proceeds as follows. Section 2 deals with unilingual and mixed DPs in Basque and Spanish (Section 2.1) and gives a brief overview of the main gender assignment strategies (GAS) reported in the literature on mixed DPs predominantly in Spanish and Basque (Section 2.2). After presenting the aims and research questions (Section 2.3), in section 3 we present the methodological details of the study: participants (Section 3.1), procedure (Section 3.2), materials (Section 3.3) and predictions (3.4). Section 4 analyses the results, which are discussed in Section 5.

2 Gender assignment

In general, (monolingual and bilingual) Spanish speakers know that some nouns like *cañón* ‘canyon’ and *mapa* ‘map’ are M(asculine), whereas *acción* ‘action’, *can- ción* ‘song’, *hierba* ‘grass’ or *mano* ‘hand’ are F(eminine). The knowledge of the morphosyntactic, phonological and semantic properties of N_S allows speakers to combine each of them with the corresponding M/F determiner –*el* ‘the_M’, *un* ‘a_M’, *este* ‘this_M’ or *la* ‘the_F’, *una* ‘a_F’, *esta* ‘this_F’ – in unilingual Spanish DPs (Corbett 2006, Roca 2005).

Each N_S has its gender specificity, which is (highly) predictable through noun endings, and less idiosyncratic than in other languages such as French or German. For instance, virtually all N_S ending in *-o* are M (99.87%) and the majority of the N_S ending in *-a* (96.30%) are F (Eddington 2002, Teschner & Russell 1984).

However, not all noun endings in this language are so “informative”, and hence the distinction between the so-called “regular”, “canonical” or “transparent” nouns (5-6) –M nouns ending in *-o* and *-a* ending F nouns– as opposed to “irregular”, “non-canonical” or “opaque” nouns (3-4) (Caffarra & Barber 2015, Vigliocco et al. 1997). Note that apart from *-a* and *-o* endings, most of other endings tend to appear in M nouns, since only 6.93% of *-i* ending, 4.9% of *-u* ending, 10.65% of *-e* ending and 51.6% of *-n* ending N_S are F.

- (3) a. el estuche
 the_M case_M
 ‘the pencil case’

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

- b. la clase
the_F classroom_F
‘the classroom’
- (4) a. la mano
the_F hand_F
‘the hand’
- b. el pijama
the_M pyjamas_M
‘the pyjamas’

Spanish grammar requires that (monolingual and bilingual) speakers “attribute” one of the two gender values available in Spanish, M or F, to the inserted noun, which will determine further combinations of the noun with many other gender-inflected words such as determiners (3-4), quantifiers (*algún/a* ‘some’; *poco/a* ‘(a)few’), and adjectives (*blanco/a* ‘white’), as well as the gender marking of other phrasal elements such as pronouns and copula-constructions.

The specificities of the languages involved in a mixed DP may affect differently bilinguals’ decisions regarding gender assignment to the inserted noun, depending on whether both languages distinguish gender (e.g. Italian and Spanish distinguishing M/F; or Sp (M/F)-German (M/F/Neuter)) or whether only one language distinguishes gender (e.g. Spanish (M/F)-English, Spanish-Papiamentu or Spanish-Basque). See § 2.2 and Bellamy and Parafita Couto (2022) for more details on the patterns found across language dyad combinations and code-switcher profiles.

2.1 Unilingual and mixed DPs in Basque and in Spanish

Basque is a language without gender marking and, consequently, the M/F opposition found in many lexical pairs in Spanish, related (5a,b) or not related to biological sex (6a,b), is realized in Basque as differentiated lexical items (5c-d, 6c-d).

- (5) a. niñ-o
child-M
- b. niñ-a
child-F
- c. . mutil
‘boy’

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

- d. neska
‘girl’
- (6) a. cerez-o
cherry-M
- b. cerez-a
cherry-F
- c. gerezi-ondo
‘cherry-tree’
- d. gerezi
‘cherry (fruit)’

The existence / absence of grammatical gender has consequences in the features of DPs in both languages, which differ in a set of additional morphosyntactic features (Corbett 2006). First, determiners and quantifiers always precede nominal elements in Spanish (7a-c), whereas variation in placement is attested in Basque, where the determiner may be the suffixed article *-a* (8a, b), a free word which follows the noun (demonstrative in (8c)) or a free morpheme preceding it (numerals (8c)). Second, determiners, quantifiers and adjectives match in gender and number with the modified noun in Spanish (7a, b, c). In contrast, no DP internal agreement is found in Basque between the noun and the rest of DP components, where all DPs’ number and case markers are attached to the last word of the DP. This last word can be a noun (8a), an adjective (8b) or a demonstrative (8c).

- (7) a. la rama b. las rama-s blanc-a-s
the.F branch the.F-PL branch-PL white-F-PL
‘the branch’ ‘the white branches’
- b. aquella-s tres rama-s
those-F-PL three branch-PL
‘those three branches’
- (8) a. adarr-a
branch-the
‘(the) branch’
- b. adar txuri-a-k
branch white-the-PL
‘(the) white branches’

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

- c. hiru adar haiek ‘those three branches’
three branch those

Noteworthy, *-a* ending nouns are the most common option when naming/listing objects in Basque. Thus, the expected response to the question *Nola esaten da “branch” euskaraz?* ‘How do you say “rama” in Basque?’ is *adarra* (8a) rather than the root **adar*. Therefore, more than one phonological form can be predicted for the lexical insertion *adar(ra)* ‘branch’ into Spanish (9), which could be parsed either as a mono-morphemic (9a) or as a complex word containing a lexical root plus a grammatical ending (9b). Similarly, the N_S *rama* ‘branch’ could be treated by the switcher as a single lexical unit (10a) or as a morphological complex (10b), made up of the root *ram* and the *-a* ending when inserting it in a Basque sentence.

- (9) a. *adarra*
b. *adarr-a*
- (10) a. *rama*
b. *ram-a*

Spanish grammar requires that every determiner matches in M/F features with the noun following it. Such a rule applies indistinctively to Spanish nouns, to borrowings and to lexical insertions in which Spanish is the host language. In order to fulfil the Det-N feature agreement requirements of the Spanish grammar on DPs, Basque nouns inserted in Spanish DPs undergo a morpho(phono)logical adaptation process of genderization, which consists in the attribution of M or F gender features to originally ungendered nouns (López 2020). Thus, a number of options are potential candidates for the producer of mixed Det_SN_B DPs containing a N_B such as *adar* ‘branch’ (11a-d).

- (11) a. *la adar*
b. *el adar*
c. *la adarra*
d. *el adarra*

2.2 Gender assignment strategies in mixed DPs

Some studies on mixed DPs provide evidence for the *analogical criterion* in mixed English-Spanish DPs, in which the gender-marked Spanish article (M)asculine

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

el or (F)eminine *la*, preceding the ungendered English N, is selected, based on the gender specificity (M or F) of the Spanish translation of the English noun (see Table 1). This analogical strategy, which has been considered the “correct” gender assignment strategy (GAS) by some researchers (Jorschick et al. 2011), is closely related to the term *translation equivalent* (Bellamy & Parafita Couto 2022), although some of the levels of similarity (e.g. morphological analogy) covered by *translation equivalent*, are not necessarily covered by the former in CS literature. In contrast, the *masculine default*, according to which speakers tend to select/prefer masculine mixed DPs, regardless of the features of the English noun inserted, is the most frequent strategy in mixed English-Spanish DPs (Otheguy & Lapidus 2003), especially for code-switchers who are dominant in the non-gendered language, English (Beatty-Martínez & Dussias 2019, Licerias et al. 2008, Valdés Kroff 2016). See also Bellamy & Parafita Couto (2022) for a more detailed discussion. Finally, the *phonological criterion* establishes the gender value of the N according to formal properties in some languages. This may be the case for Spanish, where the fact/rule that most *-a/-o* ending N_S are F/M (Teschner & Russell 1984) can be extended to lexical insertions from other (gendered or ungendered) languages.

Table 1: Examples of four gender assignment strategies in S-E and S-B mixed DPs.

	Spanish-English DPs		Spanish-Basque DPs	
	Det _S N _E	Det _S N _E	Det _S N _B	Det _S N _B
Analogical criterion	la skirt	el eye	la gona	el begi(a)
Masculine default	el skirt	el eye	el gona	el begi(a)
Feminine default	la skirt	la eye	la gona	la begi(a)
Phonological criterion	el skirt	el eye	la gona	el begi/ la begia

Diverging GAS have also been reported for mixed Spanish-Basque DPs in acceptability judgment studies, where the *masculine default* found by Badiola & Sande (2018) contrasts with the preference for the *analogical criterion* (Iriondo 2017) and also with the preference for feminine attested by Parafita Couto, Munarriz, et al. (2015). More recently, Munarriz-Ibarrola et al. (2022) discussed the masculine default and the feminine default as hypotheses compatible with the *phonological criterion* (4). This criterion, which establishes the gender value of the Det_S in mixed Spanish-Basque DPs according to the N_B word ending, is compatible with the tendency to assign F to lexical insertions with *-a* ending (pronounced as /a/ in production and written as “a” in acceptance tasks) and to assign

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

M to words with *-o*, other vowels and consonantal endings. Furthermore, elicitation data analyzed by Munarriz-Ibarrola et al. (2022) confirmed the existence of different GAS among bilingual users of the same language pair, in line with Bellamy et al.'s (2018) study on mixed Purepecha-Spanish DPs.

Interestingly, the differences in the production patterns identified among the Basque-Spanish bilinguals tested, all of them regular users of both languages, were related to the participants' L1. That is, L1S participants, who were either successive bilinguals with Spanish as their only L1 or simultaneous bilinguals with Spanish as one of their first languages, relied on the analogical criterion rather than on the phonological one. In contrast, successive bilinguals with Basque as their only L1 (and Spanish as their L2 or L2S) often relied on the phonological ending of the word, as evidenced by their preference for F determiners preceding *-a* ending nouns.

Note that Badiola & Sande (2018) identified two opposed gender preferences: a generalized preference for M Det *el* over F Det *la* with N_B like *bikote* 'couple' *batzar/ batzarra* 'meeting'; and a F preference with a specific subclass of *-a* ending words such as *la makila* 'the stick'. In this subclass the *-a* ending of the word matches the root ending *-a*. Thus, the phonological GAS ("choose F for *-a* ending roots/words, M for the rest"), seems to provide a more satisfactory explanation than the F default.

Summing up this section, gender features in Spanish are item-specific abstract values mostly associated to nouns' lexical and formal features (animacy, word ending, a.o.), which can be overtly expressed as M/F/(Neuter) marks in N_S and other categories agreeing with them (determiners, adjectives, quantifiers). Both the specific gender value (overtly expressed or not) associated with each N_S and the Det-N agreement rule inside DPs are stable components across varieties of Spanish. There is enough evidence in the research on Spanish bilingualism consistent with the assumption that similarly to unilingual DPs, where Det_S share obligatorily gender features with the N_S they precede, Det_S concord with the inserted N in mixed DPs. In contrast, the variation found regarding the overt gender features of mixed DPs, in and across bilingual communities, is compatible with the (co-)existence of different GAS in the use of lexical insertions, which can be modulated by speakers' L1, input and task.

2.3 Aims and research questions

The current paper aims to respond to two main research questions (RQ). RQ1 is related to the gender feature assignment to novel/inserted nouns in the host language Spanish, as well as to the gender feature agreement between Det_S and N_B

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

in Basque-Spanish mixed DPs. More specifically, RQ1 aims to test whether any of the four GAS (1-4) reported in previous studies prevails as the best account of the acceptance patterns of Basque-Spanish mixed DPs (Det_SN_B) by early bilinguals with high functionality in the use of those two languages:

1. Phonological GAS, according to which inanimate N_B with *-a* ending are assigned F gender whilst inanimate N_B with *-o*, as well as most N_B with *-i*, *-e*, *-u*, *-n* and *-tz* endings, are assigned M gender.
2. Analogical GAS, according to which inanimate N_B are assigned the gender feature of the Spanish translation equivalent.
3. Masculine default strategy, according to which inanimate N_B tend to be assigned M gender.
4. Feminine default strategy, according to which inanimate N_B tend to be assigned F gender.

RQ2 is whether differentiated patterns associated with bilingual profiles can be identified in acceptability tasks, in line with previous experimental studies on mixed DPs controlling for word ending and analogical gender of the inserted N_B. Munarriz-Ibarrola et al. (2022) demonstrated that cued GAS rather than (M- or F-) default strategies account for the patterns found in the elicited production of mixed Basque-Spanish DPs and that GAS variation is related to bilinguals' L1, in the sense that:

- i. Basque-Spanish bilinguals with Spanish as their first (L1S) and dominant language appear to be more strongly attuned to the analogical strategy than to the rest of GAS, whilst
- ii. Basque-Spanish bilinguals with Spanish as second language (L2S) and Basque as their first and regularly used language follow the phonological GAS predominantly.

3 Method

3.1 Participants

Twenty-three young Spanish-Basque bilinguals participated in this study. Participants that rated 4 or more of 10 quality control unilingual sentences as incorrect

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

were excluded from the study. One participant was excluded for this reason. The final sample was based on 22 participants (mean age 26.9, SD 4.9). Participants were divided into two groups based on bilinguals' language dominance: 11 *L2 Spanish* bilinguals (L2S) and 11 *L1 Spanish* bilinguals (L1S) (including both simultaneous bilinguals (2L1) and L1 Spanish L2 Basque early sequential bilinguals (L2S)). Dominance is here understood as a function of a variety of categories such as (the language of most) exposure, and (higher) proficiency and usage, in line with Meisel (2019) and Montrul (2016). Accordingly, L2S bilinguals were all early sequential bilinguals with Basque as L1 and Spanish as L2, whilst L1S bilinguals were either early sequential bilinguals with Spanish as L1 and Basque as L2 ($n = 7$) or bilinguals exposed to and users of the two languages, Spanish and Basque, from birth (2L1, $n = 4$).

As Table 2 shows, most L2 participants (9 out of 11) live in sociolinguistic environments with over 50% Basque-speaking population ($n=2 / 8 / 1$ participants live in sociolinguistic areas (S areas) 2/3/4 respectively), whereas all L1 Spanish participants live in S areas with less than 50% of Basque speakers ($n=4 / 3 / 4$ participants live in S areas 1, 2 and 3, respectively. See *Soziolinguistika Klusterra* (2014).

Based on the information provided in the language background questionnaire (Munarriz-Ibarrola et al. 2022), all participants attended schools where Basque is the language of instruction, use both languages on a daily basis, and can be considered as highly proficient bilinguals, considering that most of them (19, 86.3%) have obtained the highest proficiency diploma in Basque (C1 or similar according to the CEFR).

Regarding self-rated proficiency and dominance, L2S participants rated themselves higher in Basque than in Spanish, as opposed to L1S participants, who rated themselves higher in Spanish. Thus, L1 and language dominance in context and use converge in this sample, where most L1S participants have Spanish as L1 (solely or together with Basque), live in Spanish-dominant areas and reported being dominant in Spanish. In contrast L2S participants rate themselves as more dominant in Basque than in Spanish, and most of them have Basque as their only first language and/or live in areas with higher use of Basque.

As for the use of codeswitching, the vast majority ($n= 20$ out of 22) acknowledged the use of CS, despite differences in frequency (mean 2.9, SD 1 in a 4-point scale, where 4 corresponds to “usually” and 1 to “never”). Only 2 participants that indicated they never use CS were L2S participants.

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

Table 2: Main characteristics of the participants by group

		L2S	L1S	Overall
		participants	participants	
N		11	11	22
Age range		21-30	23-39	21-39
Age: mean		25.36 (2,98)	28.36 (5.99)	26.86 (4.9)
(SD)				
AoA	Basque	< 2 (n = 11)	< 2 (n = 4)	< 2 (n = 15)
			< 4 (n = 7)	< 4 (n = 7)
	Spanish	< 4 (n = 3)	< 2 (n = 11)	< 2 (n = 11)
		< 6 (n = 8)		< 4 (n = 3)
				< 6 (n = 8)
Sociolinguistic	B speakers >	3 (n = 8)	3 (n = 4)	13
environment	50% (S areas	4 (n = 1)		
	3-4)			
	B speakers <	2 (n = 2)	1 (n = 4)	9
	50% (S areas		2 (n = 3)	
	1-2)			
L dominant	Basque	10	1	11
	Spanish &	1	5	6
	Basque			
	Spanish	-	5	5
Proficiency		11	8	19
diploma				
(Basque)				
Self-rated	in B	3	2.73	2.86
proficiency				
(1-3 scale)				
	in S	2.82	3	2.91
Use of CS (1-		2.72 (SD 1.2)	3.1 (SD 0.83)	2.9
4 scale)				

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

3.2 Procedure

The present study utilizes an acceptability judgment task to gain insights into the bilinguals' preferences regarding gender assignment strategies in Spanish-Basque mixed DPs. Participants carried out the acceptability task on a tablet. They were presented with auditory sentences, as in (12), and were asked to rate those sentences based on a 5-point Likert scale from 1. *bad* to 5. *good* on an answer-sheet. Audio stimuli were played automatically one after the other, and participants could hear each sentence more than once if they needed to.

Participants received the following instruction in Basque: "Let's start with the questionnaire. Please evaluate each sentence (12) on the answer sheet provided selecting the option chosen (1. *bad*, 2. *very strange*, 3. *quite strange*, 4. *strange but acceptable* or 5. *good*). Remember that you are asked to evaluate the sentence you listen to."

- (12) Ha-n vendido [la *etxe-a*] que compra-ron en el centro del
 Have-3PL sold DET house-DET_B that bought-3PL in the centre of
 pueblo
 town
 'They have sold [the_F house] that they bought in the town center'

The procedures of this study were approved by the Ethics committee of the UPV/EHU (M10/2022/095) and participants gave informed consent before they started the experiment.

3.3 Materials

A total of 95 sentences were presented in the task: 3 warm up sentences followed by a test battery. The 92-item test battery was made up of 40 experimental items containing mixed Basque-Spanish DPs controlled for length (8-12 words each) and 52 additional fillers, which included 10 unilingual control items and 42 mixed sentences.

As for the critical sentences, all the words in the sentence were in Spanish except for the mixed DP containing the N_B preceded by a Det_S, as in (13).

- (13) Busca-n a un técnico que sep-a chino para [el *irratil*]
 Look-3PL for a technician that speak-3SG Chinese for the radio
 'They are looking for a technician that speaks Chinese for [the_M radio]'

The 40 nouns for the experimental sentences were selected following several criteria: nouns with biological gender were avoided, and all the nouns had 2-3

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

syllable length (some of them were used in other CS studies, such as Munarriz-Ibarrola et al. 2022, Parafita Couto, Munarriz, et al. 2015). Moreover, as displayed in Table 3, they were coded according to: 1) the analogical gender, that is, the gender of the Spanish translation equivalent or (half masculine, half feminine); and to 2) the phonological form, related to word ending (8 items with lexical *-a*, 8 with *-a* determiner and 4 more items for each of the 6 additional phonological endings).

Table 3: Nouns used in the experimental items.

			Analogical gender	
			Masculine in S	Feminine in S
Phonological ending	<i>-a</i> ending (n=16)	Lexical (ending 1)	zuhaixka ‘bush’	ezkontza ‘wedding’
			zozketa ‘raffle’	tipula ‘onion’
			makila ‘stick’	sehaska ‘cradle’
			tximista ‘lightning’	kaiola ‘cage’
	Determiner <i>-a</i> (ending 2)	<i>-a</i> (ending 2)	baso-a ‘the forest’	intxaurr-a ‘the walnut’
			txori-a ‘the bird’	esku-a ‘the hand’
			piperr-a ‘the peper’	sudurr-a ‘the nose’
			ogi-a ‘the bread’	etxe-a ‘the house’
	Non <i>-a</i> ending (n=24)	<i>-o</i> (ending 3)	soineko ‘dress’	pospolo ‘match’
			gerriko ‘belt’	gorringo ‘yolk’
		<i>-i</i> (ending 4)	eguzki ‘sun’	irradi ‘radio’
			gosari ‘breakfast’	gerezi ‘cherry’
		<i>-u</i> (ending 4)	jostailu ‘toy’	igeltsu ‘plaster’
			zigilu ‘stamp’	otordu ‘lunch’
		<i>-e</i> (ending 4)	mendate ‘pass’	landare ‘plant’
			kafesne ‘latte’	sukalde ‘kitchen’
		<i>-n</i> (ending 4)	hegazkin ‘plane’	itogin ‘leak’
			eraikin ‘building’	babarrun ‘bean’
		<i>-tz</i> (ending 4)	berakatz ‘garlic’	eskutitz ‘letter’
			hiruhortz ‘trident’	jostorratz ‘needle’

As indicated in Section 2.1. and Section 2.2., word endings 1 and 2 (*-a* ending) are prototypically F and ending 3 (*-o*) is prototypically M. The rest of endings

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

included in ending 4 correspond to the rest of vocalic endings (-i, -u, -e) and consonantal ones such as -n, which also tend to appear in M words in Spanish. Finally, N_B with (consonantal) -tz ending were also included in the non-a ending group, which are pronounced as affricate /ts/ commonly mistaken for the affricate /tʃ/ by Spanish speakers. On the basis that M and F Det_s agree in gender features with the N_B they precede, 4 match or mismatch conditions are created taking into account two GAS: analogical and phonological ending. This matching may be based on the ending of the lexical insertion (-a ending F gender vs. the rest of endings -o, -e, -i, -u, -n, -tz ending M gender) and on the analogical gender (translation equivalent F or M). The combination of the match/mismatch options for the two GAS mentioned resulted in the four experimental conditions illustrated in examples (14) to (17). Ten items per condition made up the 40-item experimental battery.

Condition 1. Analogical match and phonological match (a1-p1), 10 items.

- (14) a. la ezkontza
the_F wedding (Spanish: boda_F)
'the wedding'
- b. el soineko
the_M dress (Spanish: vestido_M)
'the dress'

Condition 2. Analogical match and phonological mismatch (a1-p0) 10 items.

- (15) a. la irratia
the_F radio (Spanish: radio_F)
'the radio'
- b. el ogia
the_M bread (Spanish: bread_M)
'the bread'

Condition 3. Analogical mismatch and phonological match (a0-p1), 10 items.

- (16) a. el jostorratz
the_M needle (Spanish: needle_F)
'the needle'
- b. la tximista
the_F lightning (Spanish: rayo_M)
'the lightning'

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

Condition 4. Analogical mismatch and phonological mismatch (a0-p0), 10 items.

- (17) a. la mendate
 the_F pass (Spanish: puerto_M)
 ‘the pass’
 b. el sudurra
 the_M nose (Spanish: nariz_F)
 ‘the nose’

Next, 42 additional fillers were inserted among the experimental items. The fillers were mixed sentences in which the language switch happened outside the DP (i.e. between the subject and the predicate, or the main and the embedded clause). Finally, there were 10 quality control unilingual grammatical sentences (5 in Spanish, 5 in Basque) in order to ensure that the participants understood the task and proceeded accurately.

In order to minimize potential item and order effects, 8 different lists were prepared with the 92 auditory stimuli. Two different sets, A and B, were created, with almost the same auditory stimuli, except for the M/F determiner of the critical items, which alternated between set A (18a) to B (18b). Next, 4 different lists were obtained from each set (A1, A2, A3, A4, B1, B2, B3, B4) after (pseudo-)randomization of the item lists, so that no more than two experimental items of the same type appeared sequentially.

- (18) a. Busca-n a un técnico que sep-a chino para [el *irradi*]
 look-3PL for a technician that speak-3SG Chinese for the radio
 ‘They are looking for a technician that speaks Chinese for [the_M
 radio]’
 b. Busca-n a un técnico que sep-a chino para [la *irradi*]
 look-3PL for a technician that speak-3SG Chinese for the radio
 ‘They are looking for a technician that speaks Chinese for [the_F
 radio]’

3.4 Predictions

On the basis of the two research questions, the literature reviewed and the specificity of participants and test materials, our main prediction is that highly competent and regular users of both Spanish and Basque language will apply some cued,

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

that is, non-default gender assignment strategy. More specifically, the first prediction is that participants' Spanish grammar will be available as it is for monolingual Spanish speakers. The rules for well-formed DPs in Spanish establish that a) Det_S bears overt gender features in agreement with the M/F gender features of the N following it and that b) each noun bears some (abstract) gender features. Consequently, speakers of Spanish feel compelled to choose the specific Det_S according to the gender value (M or F) they attribute to unknown, novel and borrowed words, such as to N_B contained/inserted in $\text{Det}_S N_B$ DPs.

The second prediction is that by taking DP internal agreement for granted in competent Spanish bilinguals' grammatical knowledge, they will rely on the canonicity principle as a general strategy for assigning gender to novel nouns in Spanish (Phonological GAS), based on the experience that most *-a* ending nouns in Spanish are F, that nouns ending in *-o* are M, as is also the case of many nouns with other vocalic (*-i*, *-u*, *-e*) and consonantal endings (*-n*, *-tz*). Although there is no evidence for nouns ending in *-tz* in Spanish, it can be predicted that most inserted nouns with consonantal endings will be identified as non-F (Harris 1991, Teschner & Russell 1984). Accordingly, higher acceptance rates are expected for the phonological matching conditions a0-p1 and a1-p1 than for the phonological mismatch conditions a1-p0 and a0-p0. That is, mixed DPs with canonical F determiner (*la basoa* 'the_F forest_M', *la ezkontza* 'the_F wedding_F'), canonical M determiner (*el soineko* 'the_M dress_M', *el gorringo* 'the_M yolk_F'), and non-canonical M determiner (*el eskutitz* 'the_M letter_F', *el berakatz* 'the_M garlic_M') will be scored higher than items in the phonological mismatch conditions a1-p0 (*el basoa* 'the_M forest_M'; *la gorringo* 'the_F yolk_F') and a0-p0 (*el etxea* 'the_M house_F', *la soineko/ berakatz* 'the_F dress_M / garlic_M').

More specifically, the third prediction is that the bilingual participants, who are regular users and highly competent in Basque (in addition to Spanish), may be able to analyse and rate differently the two types of lexical insertions of words ending in *-a* distinguished in table 3. In DPs with N_B where *-a* ending is part of the root (ending 1), F mixed DPs are expected to obtain higher acceptability scores than M ones. That is, a1-p1 *la ezkontza* 'the_F wedding_F' and a0-p1 *la zozketa* 'the_F raffle_M' will be scored higher than *el ezkontza/ zozketa* 'the_M wedding_F / raffle_M' respectively, in line with Badiola & Sande (2018) and Munarriz-Ibarrola et al. (2022) (see also Bellamy et al. (2018) for the effect of *-a* ending words in Purepecha). However, F DPs with morphologically complex *-a* ending N_B (ending 2), that is, the ones made by the Basque root plus the grammatical suffixed article *-a*, (e.g. *la txori-a* 'the_F bird_M') may obtain lower scores. In this case, the prediction is that since bilinguals may apply the canonical gender strategy (Phonological GAS) to the root and/or to the word ending *-a*, M Det-root-gender match-

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

ing mixed DPs (a0-p0 *el etxea* ‘the_M house_F’ and a1-p0 *el baso-a* ‘the_M forest_M’) will obtain higher acceptability scores than F Det-root mismatching ones (a1-p1 *la etxe-a* ‘the_F house_F’, a0-p1 *la baso-a* ‘the_F forest_M’).

Moreover, since our bilingual participants are competent in both languages, and know the gender features of the lexical Spanish equivalent of each N_B included in the experimental sentences, the fourth prediction is that they may apply the analogical GAS. In line with what has been considered the “right” gender assignment in mixed DPs in other language contact situations (Jorschick et al. 2011), they will prefer mixed DPs in which the determiner matches in gender with the Spanish equivalent’s gender feature (analogical match, a1-p0 *el basoa* ‘the_M forest_M’, *la gorringo* ‘the_F yolk_F’ and a1-p1 *la ezkontza* ‘the_F wedding_F’, *el soineko* ‘the_M dress_M’) over mixed DPs with analogical mismatch (a0-p1 *el gorringo* ‘the_M yolk_F’, *la basoa* ‘the_F forest_M’ and a0-p0 *el etxea* ‘the_M house_F’, *la soineko* ‘the_F dress_M’).

A fifth prediction can be made regarding the possibility of both the phonological and the analogical GAS being active in competent bilinguals. In that case, the highest acceptability scores will be for a1-p1 and the lowest will be for a0-p0, whilst a1-p0 and a0-p1 conditions are predicted to be in between.

The sixth prediction is that, in the case of both GAS being available, the preference ranking among the different GAS may vary even across profiles of highly functional bilinguals. Extending the tendencies observed in elicited production to acceptability of mixed DPs (Munarriz-Ibarrola et al. 2022), it can be predicted that bilinguals with Spanish as their first and more used language (L1S) will prefer mixed DPs according to analogical GAS over others (preference sequence: a1-p1 ≥ a1-p0 > a0-p1 and a0-p0). In contrast, bilinguals with Spanish as their second language (L2S) will rely more on the phonological GAS and will score items for phonologically matching conditions higher than the non-matching ones (a1-p1 ≥ a0-p1 > a1-p0 and a0-p0). Following the previous literature on Basque-Spanish bilinguals, we do not predict that our highly proficient bilinguals will assign gender based on a default strategy.

4 Results

In this section, we present data from the acceptability judgment study carried out with Basque-Spanish early bilinguals in which participants rated one by one a list of recorded utterances. Data of 22 out of the initially 23 highly functional bilingual participants were included in the study. All these participants provided valid responses for the 220 unilingual control items. They also did so for (virtually) all the experimental ones (878 out of 880 experimental items, resulting from

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

40 items x 22 participants), by selecting one out of the scores of the 5-point Likert scale for each one: bad / very strange / quite strange / strange but acceptable / good, represented as 1 / 2 / 3 / 4 / 5 in Figure 1).

All unilingual experimental items were rated as 5 *good*. For the mixed ones, the most frequent response was *strange but acceptable* or point 4 (44.5%), followed by *quite strange* or point 3 (21%), *good* or point 5 (17%) and *very strange* or 2 (12.5%). The lowest score, *bad* or point 1, was the least frequent response (5.5%). Thus, the low frequency of 2 and 1 scores (<20% together) and the considerably high mean (3.548) and median (4) values indicate a general acceptance of mixed DPs.

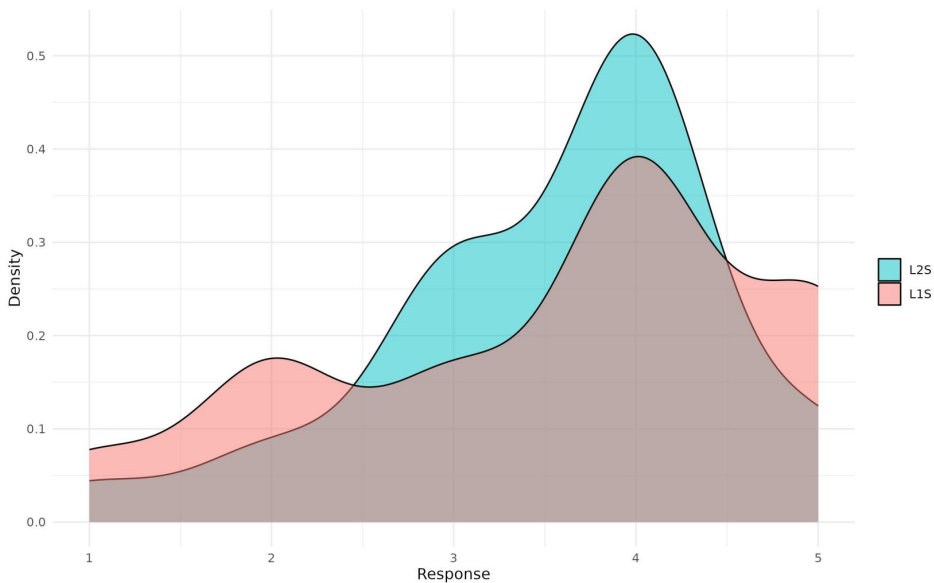


Figure 1: Distribution of the overall responses for mixed DPs per bilingual group. Note that the y-axis represents the probability density estimation, ranging from 0 to 1. This roughly translates into actual response percentages, a probability of 0.5 reflecting 50% of responses.

In general, both bilingual groups seemed to behave quite similarly with regard to the acceptability rates, as indicated by the high mean and median scores (mean: 3.54/3.55 and median 4/4 for the L1S/L2S groups respectively). Moreover, similar intragroup variability was attested, as revealed by the wide range of scores <1-5> in both groups, and the standard deviation (1.22/0.92) respectively. In fact, some participants (2 L1S and 4 L2S participants) never selected the *good* option in the scoring of mixed experimental sentences, which contrasts with their high scores for unilingual sentences (5 scores).

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

However, some differences appeared between the groups. As plotted in Figure 1, a primary peak at score 4 or *strange but acceptable* option, and a secondary peak at score 2 or *very strange*, evidenced a bimodal distribution in the L1S group's responses, which revealed as a more polarized response pattern than the L2S group's. The latter showed a centered/unimodal distribution with a single peak around score 3 *quite strange* and score 4 *strange but acceptable*, with higher frequencies than the L1S around these two values. As we will see, this difference is not statistically robust on its own, but it does interact with other factors.

Next, we compared the acceptability rates across conditions, in order to check differences between a(analogical) or p(honological) gender assignment strategies (GAS). General acceptability rates appeared as less uniformly distributed when analysing data across conditions. The most accepted condition was condition a1-p1 (19a), with acceptability rates over 68% (*acceptable* (49.09%) plus *good* (19.55%)). In this condition there was a double match, as the gender feature of the determiner matched both the gender feature of the Spanish equivalent of the Basque noun (*analogical match* or a1), and the phonological ending of the Basque noun (*phonological match* or p1). Slightly lower acceptability rates were observed for the two match-mismatch conditions, with 64.09% acceptance rate for the analogical mismatch-phonological match condition or a0-p1 (19b) and 62.73% for the analogical match-phonological mismatch or a1-p0 condition (19c), with a very small difference of 1% between them. Finally, the double-mismatch a0-p0 condition (19d) was the worst rated, which obtained the *acceptable* or *good* rate from less than half of the participants (48.64%). See Table 4.

- (19) a. a1-p1 el soineko
 the_M dress (Spanish: vestido_M)
 'the dress'
- b. a0-p1 la gazta
 the_F cheese (Spanish: queso_M)
 'the cheese'
- c. a1-p0 la jostorratz
 the_F needle (Spanish: aguja_F)
 'the needle'
- d. a0-p0 el kaiola
 the_M cage (Spanish: jaula_F)
 'the cage'

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

Table 4: Distribution of responses 1 to 5 by condition.

Condition		Responses (%)				Mean	SD	Acceptance (r4 +r5)
a0-p0	9.09	18.18	24.09	37.73	10.91	3.24	1.13	48.64
a1-p0	5.00	13.18	18.64	48.64	14.09	3.54	1.05	62.73
a0-p1	4.09	9.55	22.28	40.91	23.18	3.70	1.05	64.09
a1-p1	4.09	7.73	19.09	49.09	19.55	3.73	0.99	68.64
Analogical condi- tions								
a1-(p0 & p1)	4.54	10.45	18.86	48.9	16.81	3.635		65.68
Phonological condi- tions (a0 & a1)- p1	4.09	8.63	20.68	45	21.4	3.7125		66.36

Grouping the four conditions allowed the comparison of results by gender assignment strategies. No clear differences were observed between the analogical (a1-p0 and a1-p1) and the phonological conditions (a0-p1 and a1-p1), where frequencies of (clear) acceptance (*acceptable* + *good* responses) were almost the same for the analogical (a1p0 + a1p1) and the phonological ones (a0p0+a1p0), 65.68% and 66.36% respectively. See Table 4.

Overall, no major differences were observed between L1S and L2S groups either (mean 61% in each group) for the two acceptance responses (4 *strange but acceptable* and 5 *good* together). In contrast, rates of rejection varied: L1S group’s rejection rates (response 2 *very strange* and 1 *bad* together) were twice as big as L2S group’s, in general (24% vs. 12%), and for all the four conditions, as displayed in Figure 2. This resembles Figure 1, where L1S participants gave more polarized responses, by making greater use of the lower end of the scale than L2S participants).

As shown in Figure 2, the following ranking of conditions can be identified, which appears as consistent across the three samples (total and both group samples): a1-p1 ≥ a0-p1 ≥ a1-p0 > a0-p0. This finding contrasts, however, with the similar acceptance frequencies obtained for the two analogical and phonological

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

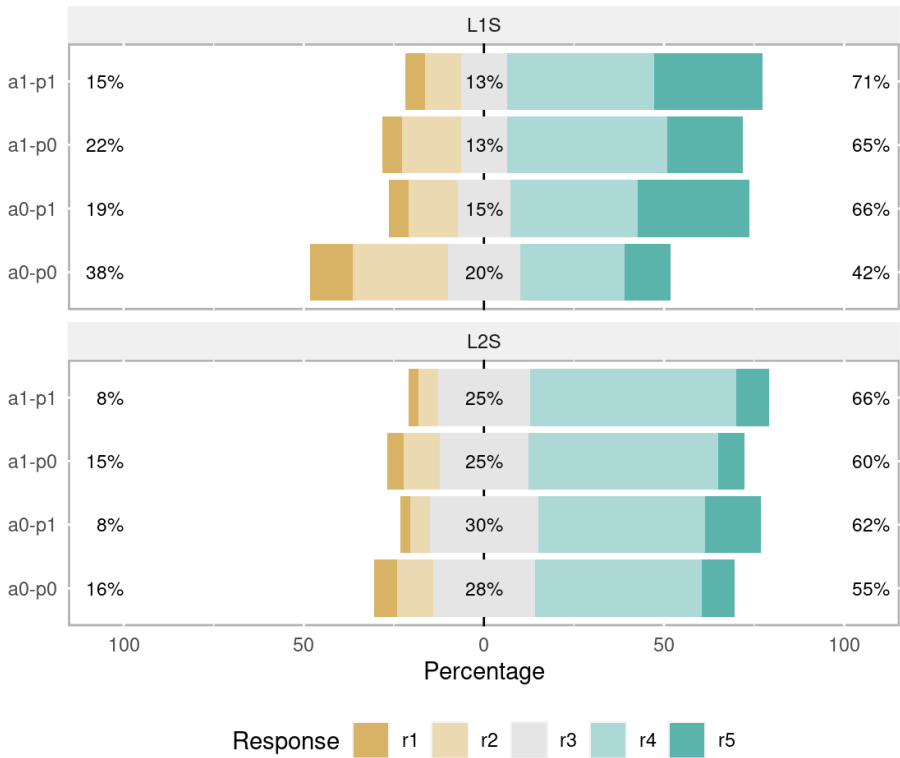


Figure 2: Distribution of responses (frequency) by condition and group

GAS in the whole sample. Figures 2 and 3b indicate that the less uniform distribution of the L1S group’s responses extended across test conditions, except for a0-p0.

Taking group and conditions into account (Figure 2), the phonology matching conditions received the highest and similar scores from both groups, with a slight difference of 4-5% between the groups, in both the additional analogical match a1-p1 (L1S/L2S 66/71%) and the mismatch a0-p1 (L1S/L2S 66/62%), followed by the a1-p0 condition (L1S/L2S 65/60%). The two groups differed the most in the worst rated mismatch condition a0-p0 (L1S/L2S 42/55%).

Noteworthy, the L1S group showed a differentiated pattern for the three (single- or double-)matching conditions – both analogical and phonological match (a1-p1), only analogical (a1-p0) and only phonological match (a0-p1) - with high acceptance rates from 65% to 71%, as compared to the considerably lower

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

acceptance rate (42%) of non-matching items (a0-p0). In contrast, the L2S group showed a more homogeneous and moderately lower acceptance rate across the three matching conditions (60-66%), and the lowest score in the mismatch condition (55%) (Figure 2).

We performed a linear mixed model (Bates et al. 2015) in order to analyse the effect as well as the interaction between the (four) experimental conditions and the two bilingual groups, L1S vs. L2S. Participant and item IDs were added as random factors to the model, in order to account for the non-independence of the responses.

Model 1: *response ~ condition * L1S + (1|participant_id) + (1|item_id)*

Model 1 revealed that experimental condition and group, taken separately, are not statistically significant predictors of judgment results (Table 5). L1S speakers showed overall lower scores, but this difference did not reach significance ($t=-1.51$, $p=0.13$). However, a noteworthy interaction arose between the two predictors (condition and bilingual group), in the sense that, compared to the L2S group, L1S speakers' responses showed statistically significant differences between conditions. Taking condition a0-p0 as a baseline, where no analogical/phonological cue is satisfied, the L1S group judged all the other conditions significantly better, the strongest effect being under condition a1-p1 ($t=3.52$, $p<0.001$), followed by a1-p0 ($t=3.07$, $p=0.002$) and a0-p1 ($t=2.69$, $p=0.007$). These effects can be visualized in Figure 2, where the L1S group gave a score of 4 or 5 to less than half (42%) of the items in condition a0-p0. This rate differs from the higher rates of 4-5 scores (66-71%) observed in the other three conditions. In contrast, the responses by L2S speakers were less polarized, and no significant differences were found between conditions.

We performed a second linear mixed model (Model 2), in order to analyse the effect as well as the interaction between the (four) experimental conditions, the M vs. F gender feature of the Spanish determiner, and the (two) bilingual Basque-Spanish groups, L1S and L2S.

Model 2: *response ~ condition * det_gender * L1 + (1|participant_id) + (1|item_id)*

Means and boxes plotted in Figure 3a indicate that acceptability rates for M vs. F differ in the range of <0.05-0.45> across conditions and groups. Here too, the L1S group (but not the L2S group) showed a differentiated pattern for matching (single- or double-) or cued and non-matching (uncued) conditions. Acceptability means were higher for M items than for F ones in the three cued conditions, where the mean differences in acceptability of M items ranged <0.08-0.45> across conditions, whereas F items scored over M in the non-matching condition a0-p0 (0.61). This suggests that the difference between a0-p0 (20a,b) and the other

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

Table 5: Results of Linear Mixed Model 1, testing the effect of experimental condition and bilingual group.

<i>Predictors</i>	<i>Estimates</i>	<i>Score</i>		
		<i>CI</i>	<i>t</i>	<i>p</i>
(Intercept)	3.43	3.03 – 3.82	17.05	<0.001***
a0-p1	0.24	-0.08 – 0.57	1.49	0.137
a1-p0	0.06	-0.26 – 0.38	0.37	0.709
a1-p1	0.21	-0.11 – 0.53	1.28	0.202
L1S	-0.39	-0.90 – 0.12	-1.51	0.130
a0-p1:L1S	0.44	0.12 – 0.76	2.69	0.007**
a1-p0:L1S	0.50	0.18 – 0.82	3.07	0.002**
a1-p1:L1S	0.57	0.25 – 0.89	3.52	<0.001***
n subj_id		22		
n item_id		80		
Observations		878		
Marginal R ² / Conditional R ²		0.042 / 0.401		

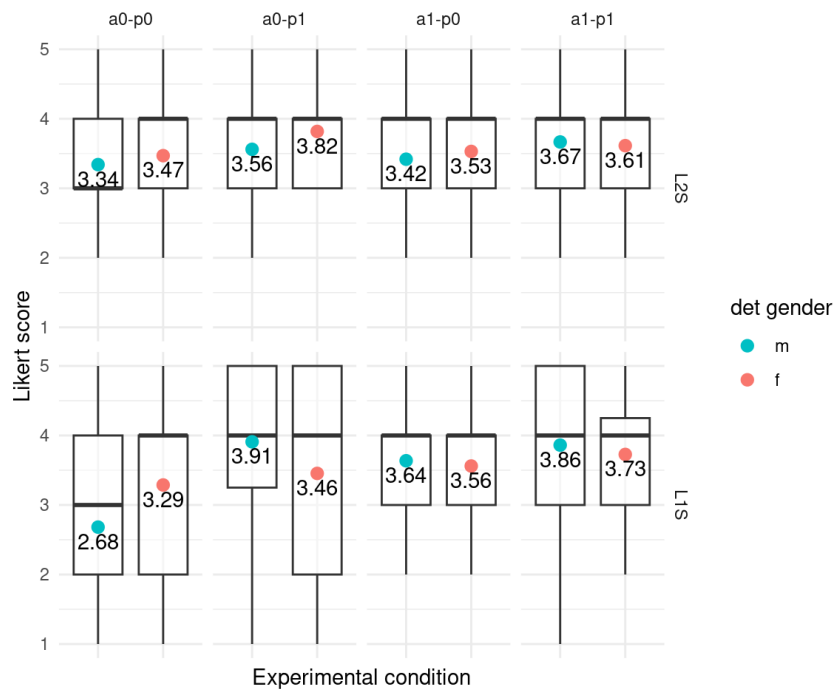
conditions observed in Figure 2 and Table 5 is mainly driven by the rejection of M Det_S preceding *-a* ending N_B under a0-p0 (20b). See also Figure 3b.

- (20) a. a0-p0 la soineko
the_F dress (Spanish: vestido_M)
‘the dress’
b. a0-p0 el kaiola
the_M cage (Spanish: jaula_F)
‘the cage’

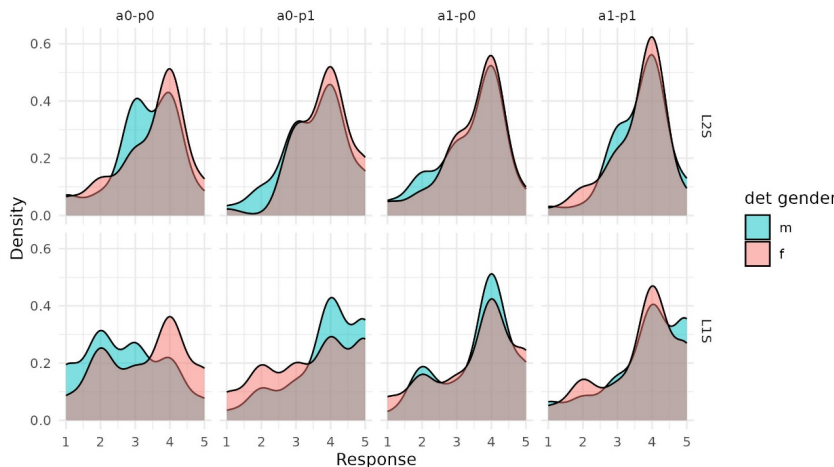
In contrast, the L2S group showed a pattern where F items were scored higher than M, in the non-matching a0-p0 condition (0.13 difference) but also in the two single matching conditions, namely a1-p0 (0.09) and a0-p1(0.26). There is one only condition in which M is scored over F, though very slightly: a1-p1 (0.06).

Taking a0-p0 condition and L2S as baseline, Model 2 revealed that gender ($t=0.53$; $p=0.594$) taken separately is not a statistically significant predictor of the judgment results (see Table 6). Nevertheless, group ($t=-2.32$; $p=0.021$) as well as the interaction of group and condition ($p<0.001$), and group and gender ($t=2.06$; $p=0.039$) revealed significant. Three-way interaction between group, condition

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola



(a) Mean acceptability rates by group, condition and gender of DetS



(b) Density plot illustrating responses by group, condition and gender of the determiner

Figure 3: Group, condition, and gender

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

and gender can be observed in the rejection of M items by the L1S group under the a0-p0 non-matching condition (20b) plotted in Figures 3a and 3b.

Table 6: Results of Linear Mixed Model 2, testing the effect of experimental condition, gender of DetS, and bilingual group.

<i>Predictors</i>	<i>Estimates</i>	<i>Score</i>		
		<i>CI</i>	<i>t</i>	<i>p</i>
(Intercept)	3.35 ***	2.87 – 3.84	13.59	<0.001
Condition 0-1	0.21	-0.25 – 0.68	0.89	0.371
Condition 1-0	0.08	-0.44 – 0.59	0.30	0.766
Condition 1-1	0.30	-0.16 – 0.77	1.28	0.202
art gender [f]	0.13	-0.34 – 0.59	0.53	0.594
L1S [1]	-0.68 *	-1.26 – -0.10	-2.32	0.021
Condition 0-1: art [f]	0.14	-0.52 – 0.80	0.43	0.670
Condition 1-0: art [f]	-0.03	-0.69 – 0.63	-0.08	0.934
Condition 1-1: art [f]	-0.17	-0.83 – 0.49	-0.51	0.607
Condition 0-1: L1S	1.02 ***	0.56 – 1.48	4.38	<0.001
Condition 1-0: L1S	0.90 ***	0.40 – 1.41	3.52	<0.001
Condition 1-1: L1S	0.90 ***	0.44 – 1.36	3.87	<0.001
art [f] × L1S [1]	0.48 *	0.02 – 0.94	2.06	0.039
Condition 0-1: art [f]: L1S	-1.22 ***	-1.87 – -0.57	-3.70	<0.001
Condition 1-0: art [f]: L1S	-0.67 *	-1.32 – -0.02	-2.02	0.043
Condition 1-1: art [f]: L1S	-0.58	-1.23 – 0.07	-1.76	0.078
n _{subj_id}		22		
n _{item_id}		80		
Observations		878		
Marginal R ² / Conditional R ²		0.059 / 0.417		
<i>p</i> <0.05 ** <i>p</i> <0.01 *** <i>p</i> <0.001				

Finally, we performed a third linear mixed model (Model 3), in order to analyse the effect of the type of *-a* ending (lexical vs. grammatical) on a subset of 88 items.

Model 3: *response ~ word ending * L1+ (1/participant_id) + (1/item_id)*

The mixed regression revealed that neither word ending (lexical vs. grammatical) nor the interaction between word ending and bilingual group reached significance (Table 7). In other words, the lexical (a1-p1 *la tipula* ‘the onion’, a0-p1 *la zozketa* ‘the raffle’) vs. grammatical nature (a1-p1 *la etxe-a* ‘the house’, a0-p1 *la soineko-a* ‘the dress’) of the *-a* ending in Basque added to the root of the inserted

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

N_B has no effect on the acceptability of the mixed DPs. Nevertheless, the different curves obtained for the two groups in the analogical mismatch conditions (a0-p0 and a0-p1) will require further research (Figure 4).

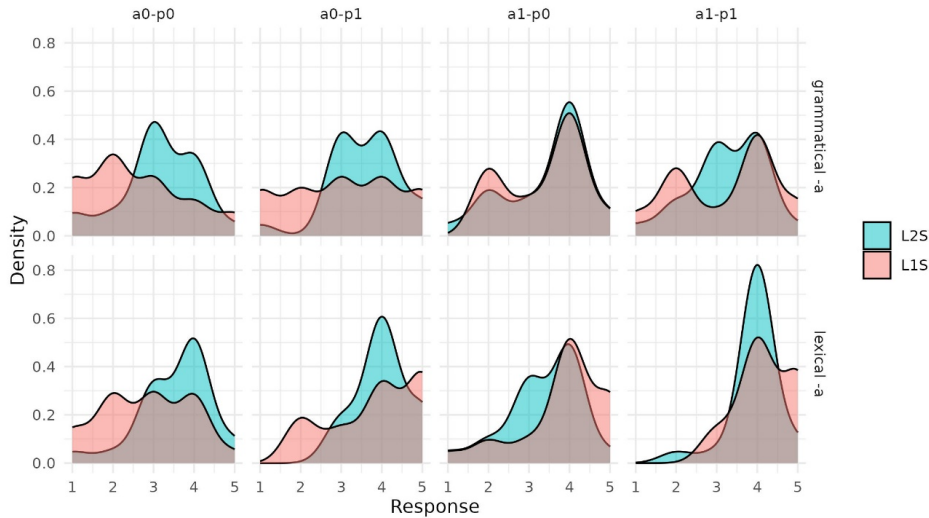


Figure 4: Density plot illustrating the acceptance responses of mixed DPs with lexical -a vs. grammatical -a, by bilingual group in four conditions.

Table 7: Results of Linear Mixed Model 3, testing the effect of -a word ending (grammatical or lexical) and bilingual group.

Predictors	Estimates	Response		p
		CI	Statistic	
(Intercept)	4.06 ***	3.40 – 4.71	12.29	<0.001
word end [2]	–0.44	–1.27 – 0.39	–1.05	0.298
spaL1 [L1S]	–0.20	–0.85 – 0.45	–0.62	0.535
word end [2] × spaL1 [L1S]	–0.40	–1.11 – 0.31	–1.12	0.266
n _{subj_id}		22		
n _{item_id}		8		
Observations		88		
Marginal R ² / Conditional R ²		0.118 / 0.471		
<i>p</i> <0.05 ** <i>p</i> <0.01 *** <i>p</i> <0.001				

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

5 Discussion and conclusions

The two groups of Basque-Spanish bilinguals tested with an acceptability task evidenced a very heterogeneous acceptance pattern for mixed sentences containing a Det_SN_B , where rates oscillated along the whole score range of 1 to 5. In contrast, they showed a consistent acceptance (rated as 5) of both unilingual Spanish and Basque sentences.

The low frequency of low scores (1. *bad* and 2. *very strange* combined (<20% together) and the considerably high mean (3.548) and median (4) values in the total sample for both L1S and L2S groups, suggests that the participants tested are familiar with mixed DPs, generally accept them, and that Det_SN_B structures in particular are part of the speech of their community.

Noticeably, 5. *good* responses were scarce and 27% of the participants (2 L1S and 4 L2S) did not select that option when judging experimental (mixed) items, which indicates that the bilinguals tested considered Det_SN_B structures *strange* or even *bad* (score 1) rather than normal, as shown by the frequency of 2-4 scores (2. *very strange*, 3. *quite strange*, 4. *strange but acceptable*). This pattern was found in both L1S and L2S groups. The individual variation attested is in line with the variation found in previous research, which tested some “conflict sites” (Poplack & Meechan 1998) in elicited and/or spontaneous production (Cruz 2023, Dubord 2004, Otheguy & Lapidus 2003, Poplack et al. 1982), but also in the acceptability of mixed sentences involving language switch in structures involving $\text{Det}_S\text{-N}_{\text{Language X}}$ structures (Beatty-Martínez & Dussias 2019, Bellamy & Parafita Couto 2022, Licerias & Fernández Fuertes 2018, Licerias et al. 2008, Valdés Kroff et al. 2017).

Unlike the bilinguals tested in the use of mixed DPs involving Det_SN_B (Munarriz-Ibarrola et al. 2022), the bilinguals tested in the current study did not show any clear preference for M or F mixed DPs. This contrasts with previous production and acceptance studies reporting M preference (Badiola & Sande 2018, Cruz 2023, Valdés Kroff 2016, Valdés Kroff et al. 2017, or F preference (Parafita Couto, Munarriz, et al. 2015) for Det_S preceding lexical insertions (N) from ungendered languages. By group, the L1S showed higher mean scores for M in the cued conditions, whilst an F preference was found in the non-matching (uncued) condition. In fact, either equal or very slightly higher mean scores for F were attested across conditions in the L2S group. The Linear mixed model 2, which included gender, bilingual group and response conditions, revealed no significant differences except for the a0-p0 condition in the L1S group, where F mixed DPs obtained significantly higher scores than M. Far from confirming the M default strategy, according to which Det_S in mixed DPs are predominantly

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

M (Beatty-Martínez & Dussias 2019, Valdés Kroff 2016), the current findings on $\text{Det}_\text{S}\text{N}_\text{B}$ are more compatible with a constraint default strategy, which is restricted to a) a particular type of words (that is, in uncued or non-matching conditions), rather than extensible to all nouns, or to broad semantic categories such as inanimate nouns, in line with Munarriz-Ibarrola et al. (2022) but in contrast to Cruz (2023); and to b) specific bilingual profiles (non-Spanish dominant) in line with Licerias et al. (2008) and Licerias & Fernández Fuertes (2018), rather than to specific communities defined in geographical or specific language-contact situations.

In line with prediction 1, sentences with $\text{Det}_\text{S}\text{N}_\text{B}$ structures did not show the same acceptance or rejection pattern across the four experimental conditions of a(nalogical) match (1) / mismatch (0) and p(honological) match (1) / mismatch (0): a1-p1, a1-p0, a0-p1 and a0-p0. The double-matched a1-p1 condition obtained the highest mean scores, followed by single-matching a0-p1 and a1-p0 conditions, and then by the non-matched a0-p0 condition, which obtained the lowest scores. This general pattern was attested in the total as well as in the two group samples, L1S and L2S.

The second prediction of higher acceptability rates for the phonological matching conditions (a0-p1 *la basoa, el gorringo* and a1-p1 *la ezkontza, el soineko*) than for the phonological mismatch conditions (a1-p0 *la gorringo, el basoa* and a0-p0 *el etxea, la soineko*) has been only partially confirmed, since total mean scores for a0-p1 and a1-p1 conditions were higher than for a0-p0. However, they were not notably higher than a1-p0 in both groups. The similar mean scores obtained in the analogical matching conditions (a1-p1 and a1-p0 combined) and the phonological matching conditions (a0-p1 and a1-p1 combined) discard the exclusion of either GAS: the phonological strategy (Prediction 2, Harris 1991, Teschner & Russell 1984) and the analogical strategy (Prediction 4). Rather than a competition, the high mean scores of the double matching condition suggests a summative compatibility of the two GAS (prediction 5). However, differences in mean scores found among the three cued conditions (a1-p1, a1-p0 and a0-p1) in the whole sample did not reach statistical significance. Statistical differences were found between the non-matching a0-p0 and the other three conditions, but only for the L1S group, where despite the small sample of 11 participants, the two analogical matching conditions (both a1-p1 and a1-p0) obtained the highest acceptability scores, followed by the only phonological matching condition (a0-p1) confirming predictions 5 and 6. The L2S group, equally small, provided data that did not reach statistical significance. Nevertheless, the acceptability pattern of L2S appeared as compatible with the prediction of a phonological GAS preference, where the doubly matching a1-p1 condition obtained the highest mean scores,

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

followed by the phonological matching condition (a0-p1). Nevertheless, the analogical GAS is also active in this group, as shown by the closeness of the mean scores obtained in the preceding and the a1-p0 condition. Both groups preferences seem to be aligned –although at different strength– with the analogical GAS or the “correct gender agreement” (Jorschick et al. 2011). Additional evidence for the analogical strategy over the masculine default in mixed DPs involving Spanish has been provided by Licerias & Fernández Fuertes (2018, submitted) in Spanish-English mixed DPs, across bilingual populations of Spanish-English bilinguals from Trinidad y Tobago (bilingual children with English vs. Spanish as heritage language, L2S speakers with advanced vs. intermediate level). Despite the differences found across tasks (acceptance, sentence completion), the higher acceptance of analogical match in mixed English-Spanish DPs for advanced than for less advanced L2S learners is in line with the assumption that variation in GAS is generally modulated by bilingual profile, and that the preference for the analogical strategy in Spanish-speaking bilinguals in particular may be associated with a higher level of competence or frequency of use of Spanish than of the ungendered language (B or E). See also Valdés Kroff et al. (2017).

The third prediction regarding the contrast between two types of *-a* endings seems not to be confirmed either, since participants did not show any preference for DPs with F Det preceding nouns with lexical *-a* endings, as compared to Basque nouns with the grammatical suffix article *-a*. Contrary to Badiola & Sande (2018), participants in the current study did not score F items such as a1-p1 *la ezkontza* ‘wedding’ or a0-p1 *la zozketa* ‘lottery’ higher than F a1-p1 *la etxe-a* ‘house-the’, a0-p1 *la ogi-a* ‘bread-the’ with grammatical *-a* ending. The mixed regression analyses did not reveal significant differences in this regard. The current findings are more in line with the acceptability results of Parafita Couto, Munarriz, et al. (2015) and the production data by Munarriz-Ibarrola et al. (2022), who concluded that cued GAS rather than default strategies could account for the variability attested. For instance, some participants (A7, A8, B8) showed more intraindividual variation than others by alternating in the form of the lexical insertion N_B in the mixed DP: (A7: *el ezitia* a0-p0, *el ezti* a0-p1 ‘the_M honig_F’; B8: *la eguzkia* a0-p1, *la eguzki* a0-p0 ‘the_F sun_F’; A8: *la ezitia* a1-p1, *la ezti* a1-p0 ‘the_F honig_F’). Interestingly, gender was found to vary interindividually *el/la ezti(a)* in those alternating pairs, but it was the form of the nominal insertion, *ezti* (root) vs. *eztia* (root + *-a* ending), rather than Det_S that showed intraindividual variation. These alternation patterns are in line with a quite stable individual representation of gender specificity as compared to the less stable representation of the form of the nominal insertion, according to which the Basque equivalent *eztia* of the Spanish noun *miel_F* ‘honey’ may be assigned only one gender option (M by

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

some bilinguals, F by others), but may vary in the form used by the same bilingual speaker when inserting it in mixed DPs: sometimes as a bare root (*ezti*) and others as a morphologically complex word (*ezti-a*).

All these findings are compatible with the first prediction that both rules, the compulsoriness of gender assignment to Basque nouns and of the overt Det-N agreement, also called “concord” (Liceras & Fernández Fuertes 2018) are active in functional bilinguals’ (productive and evaluative) use of Det_SN_B structures. They also point towards the conclusion that both the analogic and the phonological GAS drive the Basque-Spanish bilinguals’ acceptance as well as the production of mixed DPs. The different response patterns found in the acceptability task, according to which the L1S group appears as more sensitive than the L2S group to the cued/uncued distinction, as well as more analogically driven, support the claim that the strength or the hierarchy of GAS may be modulated by the bilingual profile (Liceras et al. 2008, Munarriz-Ibarrola et al. 2022, Valdés Kroff et al. 2017).

Production and Acceptability judgment data use seem to differ, in the strength of the specific GAS associated to the experimental conditions. The different GAS appear more clearly ranked for each group in production than in the acceptability judgments: analogical GAS as predominant for L1S vs. phonological GAS for L2S. The apparently contradictory results found in the two modalities of language use in Basque-Spanish mixed DPs are not an exception. Different GAS have been also identified in the production and acceptability judgment on mixed DPs by Spanish-Purepecha bilinguals. See Bellamy et al. (2018) and Bellamy & Parafta Couto (2022). In fact, different areas of the brain associated to executive functions such as the Anterior Cingular Cortex and the Dorsolateral Prefrontal Cortex) are involved in comprehension and production (Blanco-Elorrieta & Pylkkänen 2018). Moreover, the linearity of online spontaneous and even elicited oral speech, as compared to the less linear parsing and (offline) processing involved in acceptability judgment tasks, may predict the different factors affecting (differently) production and acceptability of mixed DPs. Note, for instance, that retrieving two words simultaneously in different languages (acceptability) is less cognitively demanding than having to inhibit the production of one, the dominant language (Blanco-Elorrieta et al. 2018).

In summary, the current acceptability data of mixed Det_SN_B structures by 22 Spanish-Basque bilinguals evidences that acceptability judgments on mixed DPs are driven by cued strategies, rather than by default ones. The present experiment, which was designed with the specific purpose of distinguishing the potential effects of the analogical and the phonological cues, has demonstrated that Basque-Spanish bilinguals, particularly those who demonstrate high competence

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

in both languages, rely on both the analogical and the phonological GAS, and in a summative manner in the acceptability task.

We are aware of the limitations of the current study, and that more research with a bigger sample is required in order to support the results in a more conclusive way. Nevertheless, the present findings provide further corroboration for bilinguals' profile exerting an effect on their GAS and use of mixed DPs (Bellamy & Parafita Couto 2022, Licerias & Fernández Fuertes submitted, Valdés Kroff et al. 2017).

Acknowledgments

We want to express our gratitude to the people and institutions who made this research possible. First, we are grateful to the participants who volunteered in the data sampling, and to Alejandra Iriondo for her relevant collaboration in the preparation of the materials and in the data collection, as well as to the Basque Government (IT-1627-22) and the Ministry of Science, Innovation and University (PID2023-148030NB-I00), which supported the ELEBILAB research group.

Abbreviations

CS	Code Switching	F	Feminine
DP	Determiner Phrase	M	Masculine
N	Noun	PL	plural
Det	Determiner	GAS	Gender Assignment Strategies
Det _S	Spanish Determiner	RQ	Research Questions
Det _E	English Determiner	L1	First Language
n	number	L2	Second Language
N _E	English Noun	C1	Proficiency Certificate
N _S	Spanish Noun	a0	analogical mismatch
N _B	Basque Noun	a1	analogical match
E	English	p0	phonological mismatch
S	Spanish	p1	phonological match
B	Basque	r	response.

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

References

- Badiola, Lucia & Ariane Sande. 2018. Gender assignment in Basque/Spanish mixed Determiner Phrases: A study of simultaneous bilinguals. In Luis López (ed.), *Code-switching – Experimental answers to theoretical questions: In honor of Kay González-Vilbazo*, 15–38. Amsterdam: John Benjamins.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48. DOI: [10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01).
- Beatty-Martínez, Anne L. & Paola E. Dussias. 2019. Revisiting masculine and feminine grammatical gender in Spanish: Linguistic, psycholinguistic, and neurolinguistic evidence. *Frontiers in Psychology* 10. 751. DOI: [10.3389/fpsyg.2019.00751](https://doi.org/10.3389/fpsyg.2019.00751).
- Bellamy, Kate & Maria del Carmen Parafita Couto. 2022. Gender assignment in mixed noun phrases: State of the art. In Dalila Ayoun (ed.), *The acquisition of gender: Crosslinguistic perspectives* (Studies in Bilingualism 63), 13–48. Amsterdam: John Benjamins. DOI: [10.1075/sibil.63.02bel](https://doi.org/10.1075/sibil.63.02bel).
- Bellamy, Kate, Maria del Carmen Parafita Couto & Hans Stadthagen-Gonzalez. 2018. Investigating gender assignment strategies in mixed purepecha–Spanish nominal constructions. *Languages* 3(3). 28. DOI: [10.3390/languages3030028](https://doi.org/10.3390/languages3030028).
- Blanco-Elorrieta, Esti, Karen Emmorey & Liina Pykkänen. 2018. Language switching decomposed through MEG and evidence from bimodal bilinguals. *Proceedings of the National Academy of Sciences* 115(39). 9708–9713. DOI: [10.1073/pnas.1809779115](https://doi.org/10.1073/pnas.1809779115).
- Blanco-Elorrieta, Esti & Liina Pykkänen. 2018. Ecological validity in bilingualism research and the bilingual advantage. *Trends in Cognitive Sciences* 22(12). 1117–1126. DOI: [10.1016/j.tics.2018.10.001](https://doi.org/10.1016/j.tics.2018.10.001).
- Blokzijl, Jeffrey, Margaret Deuchar & Maria del Carmen Parafita Couto. 2017. Determiner asymmetry in mixed nominal constructions: The role of grammatical factors in data from Miami and Nicaragua. *Languages* 2(4). 20. DOI: [10.3390/languages2040020](https://doi.org/10.3390/languages2040020).
- Bullock, Barbara E. 2009. Phonetic reflexes of code-switching. In Almeida Jacqueline Toribio & Barbara E. Bullock (eds.), *The Cambridge handbook of linguistic code-switching* (Cambridge Handbooks in Language and Linguistics), 163–181. Cambridge: Cambridge University Press.
- Caffarra, Sendy & Horacio A. Barber. 2015. Does the ending matter? The role of gender-to-ending consistency in sentence reading. *Brain Research* 1605. 83–92. DOI: [10.1016/j.brainres.2015.02.018](https://doi.org/10.1016/j.brainres.2015.02.018).

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

- Cantone, Katja Francesca & Natascha Müller. 2008. Un nase or una nase? What gender marking within switched DPs reveals about the architecture of the bilingual language faculty. *Lingua*. Formal syntactic approaches to bilingual code-switching 118(6). 810–826. DOI: [10.1016/j.lingua.2007.05.007](https://doi.org/10.1016/j.lingua.2007.05.007).
- Chan, Brian Hok-Shing. 2008. Code-switching, word order and the lexical/functional category distinction. *Lingua* 118(6). 777–809.
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Cruz, Abel. 2023. Linguistic factors modulating gender assignment in Spanish–English bilingual speech. *Bilingualism: Language and Cognition* 26(3). 580–591. DOI: [10.1017/S1366728922000839](https://doi.org/10.1017/S1366728922000839).
- Dubord, Elise M. 2004. Gender assignment to English words in the Spanish of Southern Arizona. *Divergencias* 2(2). 27–40.
- Eddington, David. 2002. Spanish gender assignment in an analogical framework. *Journal of Quantitative Linguistics* 9(1). 49–75. DOI: [10.1076/jqul.9.1.49.8482](https://doi.org/10.1076/jqul.9.1.49.8482).
- Harris, James. 1991. The exponence of gender in Spanish. *Linguistic Inquiry* 22. 27–62.
- Heredia, Roberto R. & Jeanette Altarriba. 2001. Bilingual language mixing: Why do bilinguals code-switch? *Current Directions in Psychological Science* 10(5). 164–168. DOI: [10.1111/1467-8721.00140](https://doi.org/10.1111/1467-8721.00140).
- Herring, Jon, Margaret Deuchar, Maria del Carmen Parafita Couto & M. Moro Quintanilla. 2010. I saw the madre: Evaluating predictions about codeswitched determiner noun sequences using Spanish-English and welsh-English data. *International Journal of Bilingual Education and Bilingualism* 13(5). 553–573.
- Iriondo, Alejandra. 2017. Kode-alternantzia elebidun gazteetan: Ds nahasiak. *Gogoa* 16. 25–56.
- Jake, Janice L., Carol Myers-Scotton & Steven Gross. 2002. Making a minimalist approach to codeswitching work: Adding the matrix language. *Bilingualism: Language and Cognition* 5(1). 69–91.
- Jorschick, Liane, Antje Endesfelder Quick, Dana Glasser, Elena Lieven & Michael Tomasello. 2011. German-English-speaking children’s mixed NPs with ‘correct’ agreement. *Bilingualism: Language and Cognition* 14(2). 173–183. DOI: [10.1017/S1366728910000131](https://doi.org/10.1017/S1366728910000131).
- Liceras, Juana M., Raquel Fernández Fuertes, Susana Perales, Rocío Pérez-Tattam & K. Todd Spradlin. 2008. Gender and number agreement in bilingual native and non-native grammars: A view from child and adult functional-lexical mixings. *Lingua* 118(6). 761–852. DOI: [10.1016/j.lingua.2007.05.006](https://doi.org/10.1016/j.lingua.2007.05.006).
- Liceras, Juana M. & Raquel Fernández Fuertes. 2018. *Formal constraints on code-switching: GFSH (Grammatical Features Spell-out Hypothesis) and gender valuation in spontaneous and experimental child data [conference talk]*. Paper pre-

Maria-José Ezeizabarrena, Amaia Munarriz-Ibarrola & Varun D. C. Arrazola

- sented at the Key debates in code-switching research: methodological and theoretical considerations, Leiden University. Leiden.
- Liceras, Juana M. & Raquel Fernández Fuertes. Submitted. Codeswitching as a window to the representation of formal features in the mind of the bilingual: Cognitive and linguistic constraints. In Maria del Carmen Parafita Couto & Ad Backus (eds.), *Routledge handbook of codeswitching*. New York: Routledge.
- Lipski, John M. 2016. The role of unintentional/involuntary codeswitching: Did I really say that? In Rosa E. Guzzardo Tamargo, C. M. Mazak & Maria del Carmen Parafita Couto (eds.), *Spanish-English codeswitching in the Caribbean and the US* (Issues in Hispanic and Lusophone linguistics 11), 139–168. Amsterdam: John Benjamins. DOI: [10.1075/ihll.11.06lip](https://doi.org/10.1075/ihll.11.06lip).
- López, Luis. 2020. *Bilingual grammar: Toward an integrated model*. Cambridge: Cambridge University Press.
- Meisel, Jürgen M. 2019. *Bilingual children: A guide for parents*. Cambridge: Cambridge University Press.
- Montrul, Silvina. 2016. *The acquisition of heritage languages*. Cambridge: Cambridge University Press.
- Moro Quintanilla, Monica. 2014. The semantic interpretation and syntactic distribution of determiner phrases in Spanish-English codeswitching. In Jeff MacSwan (ed.), *Grammatical theory and bilingual codeswitching*, 213–226. Cambridge: MIT Press. DOI: [10.7551/mitpress/8338.003.0013](https://doi.org/10.7551/mitpress/8338.003.0013).
- Munarriz-Ibarrola, Amaia, Maria-José Ezeizabarrena, Varun D. C. Arrazola & Maria del Carmen Parafita Couto. 2022. Gender assignment strategies and I1 effects in the elicited production of mixed Spanish-Basque DPs. *Linguistic Approaches to Bilingualism* 12(6). 778–815. DOI: [10.1075/lab.20016.mun](https://doi.org/10.1075/lab.20016.mun).
- Munarriz-Ibarrola, Amaia, Maria del Carmen Parafita Couto & Emma Vanden Wyngaerd. 2018. Methodologies for intra-sentential code-switching research. *Linguistic Approaches to Bilingualism* 8(1). 1–4. DOI: [10.1075/lab.17082.mun](https://doi.org/10.1075/lab.17082.mun).
- Otheguy, Ricardo & Naomi Lapidus. 2003. An adaptive approach to noun gender in New York contact Spanish. In Rafael Núñez-Cedeño, Luis López & Richard Cameron (eds.), *A Romance perspective on language knowledge and use* (Current Issues in Linguistic Theory 238), 209–229. Amsterdam: John Benjamins. DOI: [10.1075/cilt.238.17oth](https://doi.org/10.1075/cilt.238.17oth).
- Parafita Couto, Maria del Carmen, Margaret Deuchar & Marika Fusser. 2015. How do Welsh-English bilinguals deal with conflict? Adjective-noun order resolution. In Gerald Stell & Kofi Yakpo (eds.), *Code-switching between structural and sociolinguistic perspectives* (linguae & litterae 43), 64–84. Berlin: Mouton de Gruyter. DOI: [10.1515/9783110346879.65](https://doi.org/10.1515/9783110346879.65).

4 Cued GAS and L1 effects in the acceptance of mixed Basque-Spanish DPs

- Parafita Couto, Maria del Carmen, Miriam Greidanus Romanelli & Kate Bellamy. Forthcoming. Code-switching at the interface between language, culture, and cognition. *Lapurdum*.
- Parafita Couto, Maria del Carmen & Marianne Gullberg. 2019. Code-switching within the noun phrase: Evidence from three corpora. *International Journal of Bilingualism* 23(2). 695–714. DOI: [10.1177/1367006917729543](https://doi.org/10.1177/1367006917729543).
- Parafita Couto, Maria del Carmen, Amaia Munarriz, Irantzu Epelde, Margaret Deuchar & Beñat Oyharçabal. 2015. Gender conflict resolution in Spanish-Basque mixed DPs. *Bilingualism: Language and Cognition* 18(2). 304–323. DOI: [10.1017/S1366728916000572](https://doi.org/10.1017/S1366728916000572).
- Poplack, Shana & Marjory Meechan. 1998. Introduction: How languages fit together in codemixing. *International Journal of Bilingualism* 2(2). 127–138. DOI: [10.1177/136700699800200201](https://doi.org/10.1177/136700699800200201).
- Poplack, Shana, Alicia Pousada & David Sankoff. 1982. Competing influences on gender assignment: Variable process, stable outcome. *Lingua* 57. 1–28.
- Roca, Ignacio M. 2005. La gramática y la biología en el género del español. *Revista Española de Lingüística* 35(1). 17–44.
- Soziolinguistika Klusterra. 2014. *Euskararen datu soziolinguistikoen biltegia*. <http://www.soziolinguistika.eus/edb/index.php?erakus=aurkezpena> (22 January, 2020).
- Teschner, Richard V. & William M. Russell. 1984. The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics* 1. 115–132.
- Valdés Kroff, Jorge R., Paola E. Dussias, Chip Gerfen, Lauren Perrotti & M. Teresa Bajo. 2017. Experience with code-switching modulates the use of grammatical gender during sentence processing. *Linguistic Approaches to Bilingualism* 7(2). 163–198. DOI: [10.1075/lab.15010.val](https://doi.org/10.1075/lab.15010.val).
- Valdés Kroff, Jorge R. 2016. Mixed NPs in Spanish-English bilingual speech: Using a corpus-based approach to inform models of sentence processing. In Rosa E. Guzzardo Tamargo, Catherine M. Mazak & Maria del Carmen Parafita Couto (eds.), *Spanish-English codeswitching in the Caribbean and the US* (Issues in Hispanic and Lusophone Linguistics 11), 281–300. Amsterdam: John Benjamins. DOI: [10.1075/ihll.11.12val](https://doi.org/10.1075/ihll.11.12val).
- Vigliocco, Gabriella, Tiziana Antonini & Merrill F. Garrett. 1997. Grammatical gender is on the tip of Italian tongues. *Psychological Science* 8(4). 314–317. DOI: [10.1111/j.1467-9280.1997.tb00444.x](https://doi.org/10.1111/j.1467-9280.1997.tb00444.x).
- Xu, Qihui, Magdalena Markowska, Martin Chodorow & Ping Li. 2021. Modeling bilingual lexical processing through code-switching speech: A network science approach. *Frontiers in Psychology* 12. DOI: [10.3389/fpsyg.2021.662409](https://doi.org/10.3389/fpsyg.2021.662409).

Chapter 5

Sensitivity to codeswitching asymmetries in second language processing

Jorge R. Valdés Kroff^a, Hannah G. Treadway^a, Rosa E. Guzzardo Tamargo^b & Paola E. Dussias^c

^aUniversity of Florida ^bUniversity of Puerto Rico at Río Piedras ^cPennsylvania State University

Some Spanish-English bilingual communities in the US and Puerto Rico show a preference for codeswitched verb compounds involving the progressive auxiliary, *estar*. These switches are produced before or after the auxiliary with similar levels of frequency and acceptability. In contrast, verb phrases involving the perfect auxiliary, *haber*, show a clear dispreference for switches that occur within the verb compound. This same asymmetry is found in real-time sentence processing in early Spanish-English bilinguals and late second language (L2) speakers of English. We test 61 late L2 speakers of Spanish immersed in a region where Spanish-English bilingual speech is present to determine whether they exhibit online sensitivity to this asymmetry during online reading. Results indicate that L2 Spanish speakers demonstrate a sustained asymmetric pattern of sensitivity in later stages of reading but only if they engaged in a session where they answered comprehension questions. Another L2 Spanish group that completed forced-choice grammaticality judgments while reading the same sentence constructions did not show sensitivity to the same asymmetry. We interpret our findings to point towards an important role for experience with exposure to community-based code-switching patterns to successfully acquire and deploy during online sentence processing.

J. R. Valdés Kroff, H. G. Treadway, R. E. Guzzardo Tamargo and P. E. Dussias

1 Introduction

Bilinguals in the presence of other bilinguals may engage in codeswitching, the fluid use of both languages within bilingual speech or text. Bilinguals rarely code-switch accidentally when in the presence of monolinguals or in pragmatic contexts in which it is not appropriate, suggesting it is an intentional speech practice separable from other types of language mixing phenomena such as infrequent lexical intrusions (Gollan & Goldrick 2016) or lexical gap switching. Codeswitching is a complex bilingual speech act that is modulated by social, linguistic, and cognitive factors, and its use is systematic (Deuchar 2020, Torres Cacoullos & Travis 2018). Linguists have long sustained that the syntactic points at which bilinguals codeswitch follow grammatical constraints, although there is considerable debate on whether these constraints are universal, specific to bilingual grammars, or follow community-established patterns (Deuchar 2020, Liceras et al. 2008, López 2020, Parafita Couto & Gullberg 2019, Parafita Couto et al. 2023). Psycholinguistic studies have only more recently turned their attention to codeswitching (see Valdés Kroff et al. 2023, Van Hell 2022 for recent overviews). Within this body of work, the primary focus is on integration and switch costs. For successful integration to take place, bilinguals recruit domain-general cognitive mechanisms such as cognitive control to avoid conflicting representations across languages (Valdés Kroff & Dussias 2023). Switch costs refer to the putative processing costs when bilinguals encounter a codeswitch during sentence processing.

This study focuses on sensitivities to codeswitching processing constraints by second language (L2) Spanish speakers whose first language (L1) is English. Corpus-based work by teams such as those assembled and guided by Margaret Deuchar (Deuchar et al. 2014) are fundamental in pursuing this psycholinguistic line of work. Indeed, our research agenda is predicated on a detailed understanding of the interactional ecology of local bilingual communities. Our premise, thus, is that codeswitching does not need to occur at the same syntactic junctures across different bilingual communities of the same language pair and that sensitivity to these community established variations must be acquired through exposure and use. By investigating bilingual corpora, researchers can observe frequency of distinct codeswitching structures and how these structures are linked to structural constraints and community-based norms. Analyzing these production distributions then provides the testable hypotheses for investigating parsing preferences in online comprehension. We have termed this methodology the field-to-cognition approach and have outlined its premise and utility elsewhere (Beatty-Martínez et al. 2018, Valdés Kroff et al. 2018).

5 Sensitivity to codeswitching asymmetries in second language processing

Our goal is to extend prior psycholinguistic work that illustrates that high proficiency Spanish-English bilinguals (either early bilinguals or L2 English speakers) who regularly codeswitch are sensitive to the production asymmetries that emerge within their bilingual communities in the verbal phrase (Guzardo Tamargo et al. 2016). Moreover, prior work that utilizes acceptability judgment tasks demonstrates that late bilinguals' intuitions concerning acceptable codeswitching structures converge with early bilinguals and Spanish-dominant L2 English speakers (Giancaspro 2015, Toribio 2001); however, few studies have turned to the question of whether L2 Spanish speakers can deploy this knowledge online during real-time sentence processing. Moreover, by using the field-to-cognition approach, psycholinguistic studies can help disentangle the contributing roles of grammatical constraints and exposure to community-determined codeswitching patterns. We employ an eye-tracking-while-reading task on a well-attested asymmetry within the verbal phrase involving progressive and perfect verb compounds. Additionally, we introduce a task manipulation whereby participants either answered a yes/no comprehension question or completed a forced-choice grammaticality judgment task immediately after reading each sentence. This secondary task manipulation investigates how tasks commonly used in the experimental linguistics literature potentially modulate experimentally defined critical regions of interest during sentence processing in L2 participants.

1.1 Codeswitching constraints

Linguists adopt an important classification system for codeswitches, dividing them into inter-clausal (or inter-sentential) and intra-clausal (or intra-sentential) codeswitches. This division takes the sentential clause or the complementizer phrase (CP) as the syntactic unit of interest, with switches occurring across sentential clauses being inter-clausal (1) and those within as intra-clausal (2)¹.

- (1) *Ayer fui al supermercado*, and I bought some milk
'Yesterday, I went to the supermarket, and I bought some milk.'
- (2) *El maestro encontró el* book on the floor
'The teacher found the book on the floor.'

This classification coincides with important psycholinguistic considerations that occur at the individual- or group-level. For one, proficiency across the two

¹All examples throughout this paper will indicate Spanish elements in italics and English elements in regular typeface.

J. R. Valdés Kroff, H. G. Treadway, R. E. Guzzardo Tamargo and P. E. Dussias

languages is shown to affect the frequency and type of codeswitches that are produced, with lower proficiency bilinguals more likely to engage in inter-clausal and single word switching; alternatively, higher proficiency bilinguals are more likely to engage in more complex, intra-clausal codeswitching (Miccio et al. 2009, Poplack 1980). At the group level, codeswitching is also modulated by community use and acceptance (Parafita Couto et al. 2023). For instance, Poplack (1987) compared two groups of bilinguals in which a Romance language was in contact with English: French-English bilinguals from the Ottawa-Hull area of Canada and Spanish-English bilinguals from New York City. Despite the typological and macro-economic similarities of the two languages in contact, the two groups differed considerably in their frequency and types of codeswitching produced. The Spanish-English group were more likely to engage in more complex and varied codeswitches while the French-English group demonstrated a more restricted use of codeswitching. Thus, individuals and communities can vary in their patterns of codeswitching use.

However, within grammatical approaches to codeswitching, there is sustained debate on the class of constraints that regulate codeswitching. Classic approaches such as the Equivalence Constraint (Lipski 1978, Poplack 1980) posit the need for grammatical equivalence near the point of the codeswitch across the two languages. Across Spanish and English, one well known cross-linguistic difference is the surface word order of object pronouns with respect to an inflected verb. In English, it occurs post-verbally (3) but is pre-verbal in Spanish (4). Subsequently, equivalence-related constraints claim that codeswitches near the verb that involve an object pronoun should not occur (5).

- (3) After the dance, the boy **hugged her** in the street.
- (4) *Después del baile, el niño **la abrazó** en la calle.*
- (5) **Después del baile, el niño **la hugged** in the street.*

The 1980s and 90s proceeded with grammatical constraints mainly working within generative syntactic frameworks, such as the Government Constraint and the Functional Head Constraint, that specify where codeswitches are or are not licensed (Di Sciullo et al. 1986, Belazi et al. 1994). However, the late 90s saw the introduction of constraint-free or null approaches that critiqued these earlier accounts as necessitating language-specific or bilingual constraints that are vacuous among monolinguals (MacSwan 1999, Mahootian 1996). Instead, these approaches, which have become dominant in formal accounts, state that no additional mechanisms or language-specific constraints should be necessary to rationalize the grammar of codeswitches. The interaction of the two grammars

5 Sensitivity to codeswitching asymmetries in second language processing

is sufficient to account for codeswitches. These frameworks first prominently occurred within the premises and assumptions of the Minimalist Program and now additionally extend to exoskeletal and Distributed Morphology frameworks (Lohndal & Putnam 2024, López 2020). One final approach that has also contributed significantly to codeswitching theory is the Matrix Language Framework (Myers-Scotton 1993, Myers-Scotton & Jake 2000, 2001). This account takes inspiration directly from psycholinguistic models of speech production. At its core, the MLF claims that the participating languages in codeswitching are in an asymmetric relationship such that a matrix language sets the grammatical frame of codeswitched utterances whereas embedded languages contribute restricted elements, primarily content morphemes, to the grammatical frame.

At issue for this study is the natural tension that exists between grammars that constrain or license the syntactic junctures where codeswitches occur versus the individuals and communities who use and encounter them. While it is indisputable that grammar plays a fundamental role, less is known about how these structures are processed in real time, especially by late L2 speakers. This study thus investigates late L2 Spanish speakers' sensitivities to codeswitching asymmetries within the verb phrase that are well-documented for late L2 English and early Spanish-English bilingual speakers residing in the east coast of the US and in Puerto Rico. We focus on distributional asymmetries in codeswitching production because these asymmetries provide the appropriate testbed for investigating the relative contribution of grammatical knowledge, the role of community exposure, and the relationship between grammatical knowledge and sentence processing.

1.2 Mixed VPs in Spanish-English Codeswitching

While subject-predicate switches occurring between a lexical determiner phrase (DP) and its predicate have long been attested in the production of Spanish-English codeswitching (Poplack 1980), switches occurring at the auxiliary-VP boundary have less regular, more asymmetric production frequencies. Analyses of Spanish-English bilingual corpora from a variety of geographic locations, including New York City (Poplack 1980), Miami (Deuchar et al. 2014), and Gibraltar (Guzzardo Tamargo et al. 2016), reveal disparate distributional probabilities for switches involving the Spanish auxiliary *estar* 'be' and an English progressive participle and those occurring between the Spanish auxiliary *haber* 'have' and an English past participle. In both oral and written modalities, switches using the progressive structure are equally predisposed to occur at the participle (e.g., *los estudiantes están borrowing* the book 'the students are ...') as they are at

J. R. Valdés Kroff, H. G. Treadway, R. E. Guzzardo Tamargo and P. E. Dussias

the auxiliary (e.g., *los estudiantes* are **borrowing** the book). However, switches employing the perfect structure witness comparably restricted production patterns, as they are nearly categorically realized at the auxiliary (e.g., *los estudiantes* **have borrowed** the book) with significantly reduced occurrences arising within the auxiliary-VP complex (e.g., ?*los estudiantes han borrowed* the book ‘the students have ...’; Guzzardo Tamargo et al. 2016). These findings corroborate the long-established distributional asymmetries reported between *estar*+participle and *haber*+participle alternations taking place within the auxiliary phrase (Lip-ski 1978, 1985, Pfaff 1979, Poplack 1980).

Differing accounts focused on syntactic constraints or operations offer competing views on why a production asymmetry may arise. Early endeavours primarily reliant on grammatical and/or lexical constraints differ on the well-formedness of mixed VPs. Proposals like the Constraint on Closed-Class Items (Joshi 1985) and the Functional Head Constraint (Belazi et al. 1994) predict that language switching at the participial boundary should invariably yield an ungrammatical production, at least for modal auxiliary verbs. Other frameworks, whose codeswitching restrictions are principally informed by the notion of government, instead propose that all switches occurring between modals/auxiliaries and their respective participles are licensed (Di Sciullo et al. 1986). With the development of the Minimalist Program, more recent proposals deviate significantly from constraint-based accounts, attributing any illicit codeswitching at the junction of auxiliary and participle to conflicts at Phonetic Form due to head restructuring (MacSwan 2005, 2014) or to within-phase switching (López et al. 2017). These prior accounts make differing predictions on the grammaticality of mixed VPs, yet they make no clear distinction between auxiliaries.

In contrast, the Matrix Language Framework along with its corollary, the 4-M model (Myers-Scotton & Jake 2017), provide insights into the locus for this production asymmetry with subsequent consequences for processing of these structures. The central premise of these models is that some closed-class items (i.e., morphemes) are more resistant to codeswitching than others. One basic distinction is between early and late system morphemes, which as these terms indicate, are salient at different points in the retrieval and selection process during production planning and articulation. Early system morphemes like derivational affixes and satellites of phrasal verbs exhibit greater independence and subsequently can participate in codeswitching whereas outsider late system morphemes are a part of the grammatical frame, which is set by the matrix language (Myers-Scotton 1993, Myers-Scotton & Jake 2001). Given their properties in unilingual Spanish, *estar* and *haber* differ in their morphemic classification with *haber* being more uniform. Although both elements are polymorphemic and convey tense

5 Sensitivity to codeswitching asymmetries in second language processing

and aspect, *haber* is highly restricted to surfacing in a participial construction as a placeholder, a role associated with a bridge late system morpheme classification. In contrast, *estar* is multifunctional surfacing across different syntactic positions; it may express temporary aspect, need not be strictly adjacent to its participle in the progressive construction, and can function as a main verb in Spanish, the latter being a level of independence that patterns with an early system morpheme designation. Because of its multifunctional nature and greater independence than *haber*, these differences are likely to drive their asymmetric participation in codeswitching.

Regardless of the syntactic mechanisms that may constrain codeswitching production, comprehenders are confronted with the need to integrate input incrementally during online processing. As the comprehender encounters the Spanish auxiliary during reading or listening, the multifunctional nature of *estar* has a higher probability of codeswitching into English whereas the unifunctional *haber* exerts a stronger likelihood against codeswitching. Consequently, experimental data suggests that habitual codeswitchers are able to exploit the differential production frequencies of these two switch types to attenuate processing costs during real-time sentence processing, as posited by experience-based frameworks of sentence processing (Dell & Chang 2014, MacDonald 2013). That is, early Spanish-English bilinguals more readily integrate switches at the participle in online tasks of reading comprehension when the progressive structure is in use (Dussias 2003, Guzzardo Tamargo et al. 2016, Valdés Kroff et al. 2018). Some studies have replicated these results with late, Spanish-dominant L2 English bilinguals (Guzzardo Tamargo & Dussias 2013, Guzzardo Tamargo et al. 2016), while English-dominant L2 Spanish speakers' sensitivity to these production asymmetries has predominantly been investigated using offline measures such as judgment tasks. With these offline methods, L2 Spanish speakers have reliably demonstrated the ability to intuit a pattern of acceptability for auxiliary-VP switches which mirrors their frequency in production (Giancaspro 2015, Koronkiewicz 2018, Toribio 2001); importantly, L2 speakers' felicitous performance is shown to be independent of exposure to and/or use of codeswitching, but rather is found to be modulated by L2 proficiency. It remains unknown whether these results hold for online sentence processing, and whether exposure to codeswitching modulates real-time comprehension of codeswitched structures in L2 Spanish speakers.

1.3 Sentence processing, task type, and codeswitching

The role of secondary task type in (bilingual) language comprehension has only witnessed marginal investigation in studies of codeswitching processing. Valdés

J. R. Valdés Kroff, H. G. Treadway, R. E. Guzzardo Tamargo and P. E. Dussias

Kroff and colleagues 2018 tested early Spanish-English bilinguals' processing of the progressive and perfect auxiliary structures in codeswitching contexts during an eye-tracking-while-reading paradigm. In a departure from prior investigations of similar nature, an additional manipulation was introduced to the study; after reading each sentence, participants were instructed either to answer a comprehension question about what they had just read or to complete a forced-choice grammaticality judgment² for those same sentences. When answering comprehension questions requiring a yes/no response, Spanish-English bilinguals' processing of the critical codeswitches paralleled the asymmetrical distributional frequency of those codeswitches in production. However, when prompted to indicate the grammaticality of those same sentences, this sensitivity to the production asymmetry was no longer present, indicating the modulating role that secondary task type plays in online sentence processing.

The difference in online processing between the two secondary task types was ascribed to the metalinguistic nature of the forced-choice grammaticality judgment task, which likely invokes greater awareness of major syntactic boundaries. Because switches occurring within the auxiliary-VP complex seemingly "disrupt" syntactic boundaries which are perceptual units (Fodor & Bever 1965), they are dispreferred to switches occurring at the auxiliary (Valdés Kroff et al. 2018). In a similar vein, adults reading in their L1 globally adjust their eye movement patterns both temporally (i.e. producing longer and more fixations) and spatially (i.e. exhibiting shorter saccade amplitudes) when proofreading relative to reading for comprehension (Kaakinen & Hyönä 2010). This would seem to substantiate an account of (metalinguistic) task demands having regulatory effects on attentional resources during reading. Though not entirely analogous, these observations do align with some overarching findings from reading comprehension and translation. For example, adult readers take longer to process sentence structure when asked to recall texts than when asked to comprehend those same texts (Aaronson & Ferres 1986). Similarly, reading for translation, which demands the maintenance of linguistic form in working memory, requires more cognitive resources than reading for simple repetition or understanding both in translators and bilinguals (Macizo & Bajo 2006). It is yet uncertain whether this same phenomenon generalizes to L2 Spanish speakers as they read codeswitches, and whether their sensitivity to the distributional production frequencies of codeswitching varies in accordance with secondary task type.

²Throughout the paper we refer to our secondary task as a grammaticality judgment task instead of the often-used acceptability judgment task. We do so because of the instructions provided to the participants and the prompts that they saw. Following each sentence, participants responded to the prompt "Is this grammatical?"

2 Present study

The primary goal of the current study is to investigate L2 Spanish speakers' sensitivities to well-attested production asymmetries within the verb phrase in Spanish-English codeswitching. The target structures are the progressive and perfect verb compounds. Prior work with early Spanish-English and Spanish-dominant L2 English bilinguals show that participants exhibit an asymmetric preference for progressive switches at the participle (e.g., *está* cooking) as compared to perfect switches at the participle (**ha* cooked). Moreover, Spanish-English bilinguals integrate perfect switches more easily at the auxiliary boundary as compared to the participle boundary, while exhibiting no such difference for progressive switches. Based on these prior results, we hypothesize the following: If L2 Spanish speakers can deploy metalinguistic knowledge of codeswitching grammatical constraints during online sentence processing (something they have demonstrated in offline judgments; [Giancaspro 2015](#)), then they will demonstrate the same asymmetric patterns during online sentence processing as early and L2 English bilinguals.

A secondary, exploratory goal is to investigate the effects of commonly used secondary tasks in psycholinguistic studies on real-time processing of the target codeswitching structures. In a between-groups design, participants either completed comprehension questions or forced-choice grammaticality judgments after each sentence. If indeed secondary task goals can impact L2 speakers' online processing, then different effects should emerge across the two experiments. Specifically, if the L2 Spanish speakers behave similarly to the early bilingual and L2 English groups (e.g., [Valdés Kroff et al. 2018](#)), then the codeswitching asymmetry will emerge for the comprehension question group.

To test our hypotheses, we conducted an eye-tracking-while-reading study. Eye tracking affords an ecological real-time measure of sentence processing as participants read text in a naturalistic manner ([Valdés Kroff et al. 2018](#)). During the course of reading, the eyes are typically in one of three states: fixations, blinks (or looks away), and saccades. Readers launch planned saccades (i.e., ballistic movements) between fixations, which vary in their duration and their path. In written languages like Spanish and English, fixations proceed in a rightward forward fashion with some regressions to prior regions of the written text. For this study, we calculate two measures of interest: gaze duration and total duration. Based on researcher-defined regions of interest in text, eye fixations can be grouped together using various thresholds to indicate the timecourse of reading at various stages. Gaze duration is an early measure thought to index more automatic reading processes whereas total duration is a late measure that incor-

J. R. Valdés Kroff, H. G. Treadway, R. E. Guzzardo Tamargo and P. E. Dussias

porates regressions and is thought to reflect more controlled reading processes (Rayner 1998, Godfroid 2019).

2.1 Method and materials

2.1.1 Participants

The study and its procedure were approved by the University of Florida's Institutional Review Board. Sixty-one participants whose L1 was English and subsequently learned Spanish to an intermediate-high proficiency level provided informed consent and received course credit for their participation. Participants were recruited from 3000- and 4000-level Spanish courses at the University of Florida. The University of Florida's main campus is located in Gainesville located in North Central Florida. According to US Census bureau data reported from 2022, the city is 55% White (non-Hispanic or Latino), 30% Black, and 12% Hispanic or Latino³. The university itself reports an undergraduate population that is 24% Hispanic or Latino⁴. These demographic data only provide a snapshot based on ethnicity; however, they suggest that the late bilingual group that participated in this study has access to and potentially have interactions with Spanish-English bilinguals or Spanish-dominant speakers. Moreover, the university attracts students from South Florida, a region that is known for its highly bilingual and Spanish-dominant characteristics, with a prevalence of bilingual speech (Carter & Lynch 2015, Deuchar et al. 2014).

After providing informed consent, the participants were divided into one of two experimental sessions, a comprehension question task group ($n = 30$) or a forced choice grammaticality task group ($n = 31$). Participants sat with their heads rested in a chin rest to provide head stability, which was positioned before a desk-mounted Eyelink 1000 Plus eye tracker (SR Research) and a Benq 24" computer monitor. The computer monitor was positioned approximately 70 cm from the participants' eyes. Following brief instructions, participants' eye movements and fixations were calibrated following a 9-point procedure, where errors that averaged below 0.55° were accepted before proceeding with the main experimental task. The experiment began with 5 practice trials to familiarize the participants with the task and the presence of codeswitched text.

Following the main task, participants completed a Spanish grammar test adapted from the highest level certification of the DELE⁵, a standardized assessment tool for L2 Spanish developed by the Ministry of Education, Culture, and

³<https://data.census.gov/table?g=160XX00US1225175>

⁴<https://ir.ua.ufl.edu/facts/diversity/>

⁵The version of the DELE that we use is of a more advanced level than the adapted version of

5 Sensitivity to codeswitching asymmetries in second language processing

Sport of Spain (<http://diplomas.cervantes.es/en>), an in-house language history questionnaire (LHQ), and an in-house codeswitching background questionnaire, which we have previously used in other studies (e.g., Guzzardo Tamargo et al. 2016, Valdés Kroff et al. 2017). We provide the questions used in the questionnaire as a supplemental file on the Open Science Framework⁶.

The entire experimental procedure took less than one hour per participant. Participant background characteristics and proficiency measures are reported in Table 1. From the LHQ we asked participants to indicate in years when they first began acquiring Spanish. We also asked them to rate on a scale of 1-10 (10 is high fluency) their Spanish proficiency across four domains: speaking, listening, writing, and reading. After providing examples of codeswitching and a short definition, participants were asked to rate on a scale of 1-5 (5 is always) their own likelihood of engaging in (CS Use) or being exposed to (CS Exposure) bilingual code-switching with family and friends. Two-sample t-tests were performed for each metric and all *ps* were greater than 0.325, indicating no evidence for differences between groups on each measure. Importantly, for purposes of this study, the lack of between-group differences, especially for CS use and exposure, indicates that the groups are well matched. Subsequently, if the results reveal differences in reading patterns across experiments, they are suggestive of the secondary task inducing different processing behavior across groups.

2.1.2 Materials

The materials used in this study come from Guzzardo Tamargo et al. (2016). They include 32 experimental sentence frames and 64 filler sentences. The experimental sentence frames always began in Spanish and were complex sentences in which the embedded clause consisted of a subject NP always presented as 3rd person plural, a critical compound verb (*estar/haber* + participle), a direct object, and varying adjuncts. Experimental sentence frames were constructed such that they were controlled for various lexical and syntactic features, including sentence length and lexical frequency and verb regularity of participles. Filler trials were of similar length but differed in their syntactic structures and where codeswitches occurred (e.g., between verb and direct object, between definite

the DELE that is commonly used in other studies. We have used and reported on this version of the DELE in prior studies from our group (e.g., Dussias et al. 2013, Guzzardo Tamargo et al. 2016, Valdés Kroff et al. 2017); however, we interpret this measure relatively as an index of participant proficiency and not as a proficiency-level classifier or to draw direct comparisons to other studies that use a different adapted DELE.

⁶https://osf.io/m2ew9/?view_only=74b6fc832acd45e9858fe06cca3221d5

J. R. Valdés Kroff, H. G. Treadway, R. E Guzzardo Tamargo and P. E. Dussias

Table 1: Bilingual Participants’ Characteristics Note: Mean values are presented with standard deviation reported in parentheses. The L2 self-report measures come from the LHQ where participants were asked to rate their own proficiency on a scale from 1 to 10, with 10 indicating L1 proficiency. The CS self-report measures are from a CS background questionnaire on a scale from 1 to 5, with 5 indicating “always.” The adapted DELE test reports scores out of 50 multiple choice responses from the C2 highest level.

Measure	Comprehension Question Group	Grammaticality Judgment Group
L2 Age of Acquisition (in years)	12.6 (3.3)	12.2 (3.1)
L2 Speaking (out of 10)	6.4 (1.2)	6.1 (1.1)
L2 Listening (out of 10)	6.7 (1.1)	6.7 (1.2)
L2 Writing (out of 10)	6.7 (1.7)	7.1 (1.7)
L2 Reading (out of 10)	6.8 (1.4)	7.2 (1.7)
CS Use (out of 5)	3.3 (1)	3.2 (1)
CS Exposure (out of 5)	3.3 (0.7)	3.1 (1)
DELE (out of 50)	22.9 (4.5)	22.1 (3.3)

article and noun, or between clauses). Prior to the start of each sentence trial, participants focused on a left-aligned fixation point positioned in the same location as the beginning of a new sentence. Sentences were presented in 14 pt Consolas font. For further specific details on stimuli construction and for a full list of items, see [Guzzardo Tamargo et al. \(2016\)](#).

Experimental sentence frames were manipulated such that at the critical verb compound, participants either encountered the progressive or the perfect auxiliary and its corresponding participle in English: present participle for progressive and past participle for perfect. Additionally, the codeswitch into English either occurred before the auxiliary verb or before the participle. This 2 x 2 manipulation led to 4 unique condition types, exemplified in Table 2.

With this experimental design, the critical region always occurred in the same sentential position across all experimental trials, and preceding content was syntactically similar across all critical items. Due to its placement in the middle of the sentence, the critical region also appeared positionally in the middle of the computer screen for each trial.

Participants were randomly assigned to the Comprehension Question session or the Grammaticality Judgment session. The Comprehension group encountered a Yes-No comprehension question after every trial. Because the trials al-

5 Sensitivity to codeswitching asymmetries in second language processing

Table 2: Example experimental sentence frame and critical experimental manipulation

Sentence preamble	Auxiliary	Participle [Critical]	Direct Object [Spillover]
<i>El guía piensa</i>	are <i>están</i>	enjoying	the food ...
<i>que los turistas</i>	have <i>han</i>	enjoyed	

ways began in Spanish and switched into English, the comprehension questions were always presented in English. Half of the questions required “yes” responses via button box response. The Grammaticality Judgment group saw the prompt “Is this grammatical?” and simply responded via button press whether they thought each sentence was grammatical or not. No feedback was provided for either task manipulation.

2.1.3 Data pre-processing

The eye-tracking while reading technique provides multiple reading measures of interest based on eye movements that occur within experimentally defined regions of interest while considering their sequence and time course into and out of these regions. In this study, we report on two primary measures of interest in two regions of interest. We extracted the raw eye fixation report for all participants from the eye tracker using the automatically defined word-by-word interest areas provided by Data Viewer (SR Research). We then used the Get Reading Measures python script provided by SR Research to calculate gaze duration, defined as the sum of all first-pass fixations in a region before the eye exits the region, and total duration, defined as the sum of all fixations that occur in a region, including regressions back into that region. Gaze duration is considered an early measure of sentence processing whereas total duration is a late measure that incorporates all time spent re-reading in a region (Rayner 1998). These two measures were calculated for the critical region where participants encountered the participle (Word 8), and a spillover region which combined the following two words (i.e., Words 9-10, the direct object) as illustrated in Table 2. Since our participants were L2 readers of Spanish, the ability to tap into earlier and late reading measures at two points in the experimental sentence is of particular interest since prior work has demonstrated that L2 readers may show sensitivity to

J. R. Valdés Kroff, H. G. Treadway, R. E. Guzzardo Tamargo and P. E. Dussias

grammatical manipulations at later points in time as compared to L1 or high proficiency readers (Avery & Marsden 2019, Godfroid 2019). In preparation of our analysis, we removed participants whose data track loss was greater than 80% (one participant) and removed trials on which fixations registered less than 150 ms, which is too small of a threshold to be a meaningful fixation (Rayner 1998). Below we report the percentage of data trimmed per reading measure.

For the analysis, we used logistic and linear mixed-effects regression models using the `lme4` package (Bates et al. 2015; version 1.1-35.3) within the R language (version 4.4.1). For our critical and spillover region, we included Auxiliary (progressive, perfect), Switch Position (at auxiliary, at participle), and Task (Comprehension Question, Grammaticality Judgment), and their interactions as fixed effects. P-values were calculated using the Satterthwaite’s approximation via the `lmerTest` package (Kuznetsova et al. 2017; version 3.1-3). These categorical independent variables were all contrast coded with the following scheme: at participle = 0.5, perfect = 0.5, and comprehension question = 0.5. For the random effects structure, we adopted a theoretically motivated maximal structure, by including Auxiliary, Switch Position, and their interaction with random slopes and random intercepts on Participants and random slopes of Task and Switch Position and random intercepts for Items. If these models failed to converge, then random effects structures were simplified in a step-by-step fashion, first by removing interactions, then correlations with random intercepts, then eliminating a factor until achieving model convergence (Barr et al. 2013). In anticipating our results, we find an interaction by Task, thus motivating us to discuss the results split by group, providing greater interpretability of the main findings. In the separate by-task analyses, significant interactions were then followed by pairwise tests corrected for multiple comparisons using the Tukey’s HSD method with the package `emmeans` (Lenth 2024; version 1.10.2).

3 Results

3.1 Full analysis

While not a central focus of the analysis, we first report on the task performance results for each group on the experimental trials (32 experimental sentences total)⁷. These results are not directly comparable across groups as the nature of the tasks differ. For the Comprehension Question group, participants’ accuracy

⁷For ease of exposition, we mainly report significant findings. Full model outputs for fixed effects are provided as a supplemental file at https://osf.io/m2ew9/?view_only=74b6fc832acd45e9858fe06cca3221d5.

5 Sensitivity to codeswitching asymmetries in second language processing

was calculated by assigning a ‘1’ for a correct ‘yes’ or ‘no’ response and ‘0’ for incorrect responses. This group had a mean accuracy of 92% (SD = 11%). There was minimal variation across conditions, range = 90%–93%. To test for differential responses across conditions within this group, we conducted a mixed effects logistic regression, crossing Auxiliary (*estar*, *haber*) and Switch Position (at the auxiliary, at the participle) and including random intercepts by participants and items. No main effects or interaction were significant (all $p \geq 0.59$), indicating no evidence for different performance on comprehension question accuracy.

For the Grammaticality Judgment group, participants were asked to indicate whether the previously read sentence was grammatical or not via button press (i.e., a forced-choice grammaticality judgment task). Note that there is considerable debate on the grammaticality of these mixed verb constructions (see [Balam et al. 2020](#), [Giancaspro 2015](#), [Guzzardo Tamargo et al. 2016](#)). For convenience, we coded the experimental sentences as ‘1’ if the participants indicated grammaticality for all switches except for perfect switches that take place at the participle (*los estudiantes han borrowed...“the students have borrowed...”*), in part, motivated by the infrequent attestation of this codeswitch type in oral and written corpora ([Guzzardo Tamargo et al. 2016](#)), and low acceptability even among L2 speakers of Spanish ([Giancaspro 2015](#)) and ‘0’ otherwise. Following this coding criterion, the participants demonstrated an overall mean accuracy of 65% (SD = 35%, range = 24%–82%). This lower mean percentage is skewed by the participants’ greater “inaccuracy” on codeswitched sentences that contain a perfect auxiliary and switch at the participle—as in the illustrated example above. This finding indicates that participants were overall more willing to consider this condition grammatical in contrast to the low frequency attestation in corpus findings among high proficiency bilinguals and prior judgment data work, even among L2 speakers of Spanish ([Giancaspro 2015](#)). However, prior L2 Spanish studies using judgment data also indicate that such sensitivities increase with greater Spanish proficiency ([Giancaspro 2015](#), [Koronkiewicz 2018](#)), a manipulation not tested in the current study. After removing this specific condition, the participants show minimal variability on whether they find the remaining codeswitching conditions grammatical, mean accuracy = 79% (SD = 25%), range = 75%–82%.

Reflecting the overall distribution of percentages across conditions, the mixed-effects logistic regression model showed a significant interaction between switch position and auxiliary type ($b = -2.3$, $SE = 0.346$, $z = -6.634$, $p < 0.001$), and main effects for switch position ($b = -1.58$, $SE = 0.175$, $z = -9.06$, $p < 0.001$) and auxiliary type ($b = -1.34$, $SE = 0.175$, $z = -7.665$, $p < 0.001$). The negative coefficients mean that participants were less accurate overall on perfect trials and with switches at the participle and the interaction revealing that these effects are

J. R. Valdés Kroff, H. G. Treadway, R. E. Guzzardo Tamargo and P. E. Dussias

differential across auxiliary conditions, once more pointing towards the lower accuracy (i.e., greater indication of grammaticality) of trials in which codeswitches occur at the participle with the perfect auxiliary.

We move on to the linear mixed-effects regression models that include the variables Auxiliary, Switch Position, and Task, and their interactions. First, we present model outputs for the analyses on the critical word, the participle, which was the 8th word across all experimental trials. For gaze duration, the model indicates an auxiliary by task interaction ($b = -43.36, SE = 16.86, t = -2.04, p = 0.041$), a switch position by task interaction ($b = -43.36, SE = 16.89, t = -2.57, p = 0.011$), as well as a main effect for switch position ($b = 72.4, SE = 8.44, t = 8.58, p < 0.001$). For total duration, there are two interactions, one between switch position and auxiliary ($b = 115.38, SE = 55.03, t = 2.097, p = 0.036$) as well as an interaction between the auxiliary and task ($b = -184.92, SE = 69.07, t = -2.677, p = 0.009$). We once again observe a main effect for switch position ($b = 196.15, SE = 34.35, t = 5.71, p < 0.001$). The positive coefficients for the main effect of switch position for both reading measures indicate overall slower reading times on the participle. The interaction suggests that participants read the *estar* trials differentially as compared to the *haber* trials and these effects were conditioned by the secondary task that participants completed. We continue by reporting on analyses separated by group below and additionally analyze the immediately following spillover region (combined words 9 and 10).

3.2 Comprehension group analysis

For the critical region, the data trimming procedure resulted in a reduction of 34% of the dataset for gaze duration (the early reading measure) and 20% of the dataset for total duration. For the spillover region, this trimmed 27% of the dataset for gaze duration and 20% of the dataset for total duration.

3.2.1 Critical word region (Word 8)

For the linear mixed-effects model fit to the gaze duration dataset, the model that converged included a random effects structure with random intercepts and slopes for Auxiliary on Participants and random intercepts on Items. This model revealed a main effect for Switch Position such that switches at the participle were read more slowly than those at the auxiliary ($b = 51.03, SE = 11.49, df = 13.21, t = 4.44, p < 0.001$). Neither the effect of Auxiliary type nor the interaction was significant (Auxiliary type: $b = -1.18, SE = 12.24, df = 31.93, t = -0.01, p = 0.924$; interaction: $b = 6.26, SE = 22.95, df = 271.44, t = 0.273, p = 0.785$).

5 Sensitivity to codeswitching asymmetries in second language processing

The model that converged for the total duration dataset has a random effects structure with random intercepts and slopes for Auxiliary on Participants and random intercepts on Items. This model showed a significant interaction between Switch Position and Auxiliary Type ($b = 207.2, SE = 78.73, df = 246.43, t = 2.632, p = 0.009$) and a main effect for Switch Position ($b = 159.95, SE = 39.38, df = 245.89, t = 4.061, p < 0.001$). To investigate the interaction further, we conducted pairwise tests, which showed two significantly contrasting pairs with a critical pair of interest involving the perfect auxiliary, *haber*: switches at the participle (*los estudiantes han borrowed*) were read more slowly than those at the auxiliary (*los estudiantes have borrowed*) (difference = 263.6, $SE = 56.3, df = 245.6, t = 4.682, p < 0.001$) and one contrast of less theoretical importance, where participants read the critical region faster on switches at the participle involving the progressive structure (*los estudiantes están borrowing*) as compared to switches involving the perfect auxiliary and having switched at the auxiliary (*los estudiantes han borrowed*) (difference = 197.2, $SE = 62.3, df = 89, t = 3.164, p = 0.011$). All other contrasts were not significant (all $p > 0.113$). These results are graphically presented in Figures 1 and 2.

3.2.2 Spillover region (Words 9 and 10)

On the gaze duration dataset, the linear mixed-effects model had a random effects structure that only included random intercepts on Participants and Items. This model did not reveal any interaction or main effects (all $p > 0.48$).

However, the model fit for the total duration dataset indicates a significant finding. This model had a random effects structure that includes random intercepts and slopes for Switch Position and Auxiliary (no interaction) on Participants and random intercepts on Items. The interaction between Switch Position and Auxiliary was significant ($b = 162.6, SE = 53.73, df = 228.48, t = 3.026, p = 0.003$). None of the main effects were significant (all $p > .137$). We conducted pairwise tests on the interaction and as before, two contrast were significant: the contrast between perfect auxiliary codeswitches in which switches at the participle were significantly slower than switches at the auxiliary (difference = 112.2, $SE = 39.4, df = 102, t = 2.848, p = 0.027$) as well as the contrast between the switch at participle trials with switches involving the progressive being read faster than those involving the perfect (difference = 126.8, $SE = 40.6, df = 87.9, t = 3.121, p = 0.013$). As only the model for total duration shows a significant finding, we present a graph only for those results in Figure 3.

J. R. Valdés Kroff, H. G. Treadway, R. E. Guzzardo Tamargo and P. E. Dussias

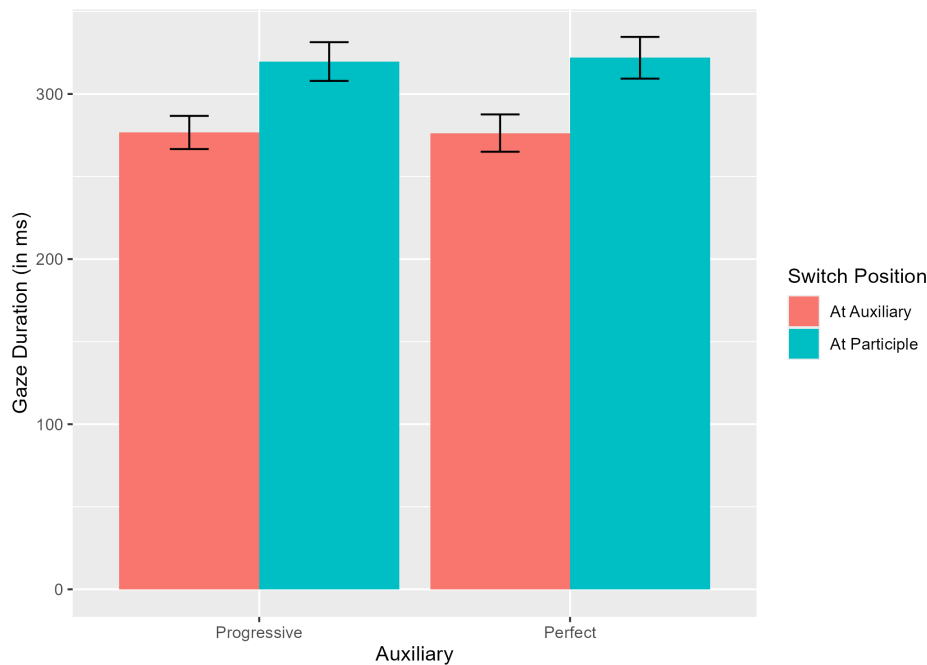


Figure 1: Mean gaze duration reading time in milliseconds on the critical region, the participle, graphed by Auxiliary Type and Switch Position. Error bars represent standard error of the mean.

To summarize, on the critical word region (i.e., the participle), L2 Spanish readers tasked with completing a comprehension question showed sensitivity to switch position in the early reading measure of gaze duration, reading switches at the participle more slowly than at the auxiliary irrespective of auxiliary. In the later reading measure of total duration, these L2 participants demonstrated a differential sensitivity to switch position based on the auxiliary. The participants showed no difference in reading times across switch position for the codeswitches that include the progressive auxiliary, *estar*. However, for codeswitches that include the perfect auxiliary, *haber*, L2 Spanish participants read the codeswitches at the participle more slowly than at the auxiliary. This finding mirrors that of early Spanish-English bilinguals and L2 English participants and is consonant with the production asymmetry found in the immediate bilingual environment of these participants. We turn now to the results for the L2 Spanish group that completed a forced-choice grammaticality judgment task.

5 Sensitivity to codeswitching asymmetries in second language processing

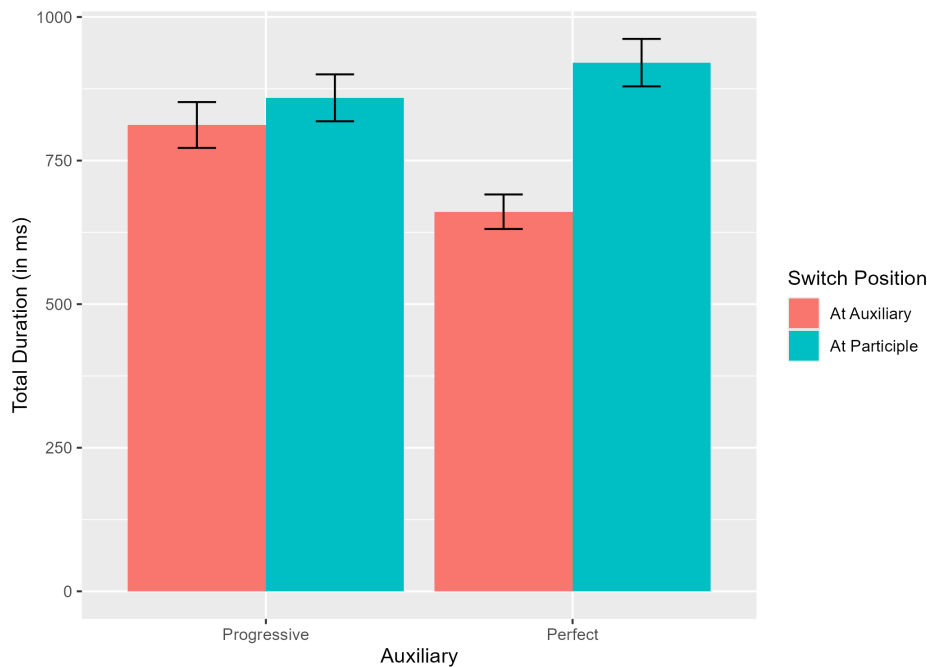


Figure 2: Mean total duration reading time in milliseconds on the critical region, the participle, graphed by Auxiliary Type and Switch Position. Error bars represent standard error of the mean.

3.3 Grammaticality judgment group analysis

For the group that completed grammaticality judgments, we only removed trials on which reading measures were below 150 ms as the grammaticality of these trials does not have a clear target, especially codeswitches with the perfect auxiliary that switch at the participle. This data trimming procedure resulted in a reduction of 56% of the dataset for gaze duration and 49% of the dataset for total duration. For the spillover region, data trimming resulted in a loss of 53% of the dataset for gaze duration and 49% for total duration. Immediately, we observe that the reading patterns for this group are considerably different than for the participants who completed the comprehension question task. This difference is not attributable to proficiency (self-reported or assessment based), age of acquisition, or immersive differences as noted by the descriptive statistics reported in Table 1.

J. R. Valdés Kroff, H. G. Treadway, R. E Guzzardo Tamargo and P. E. Dussias

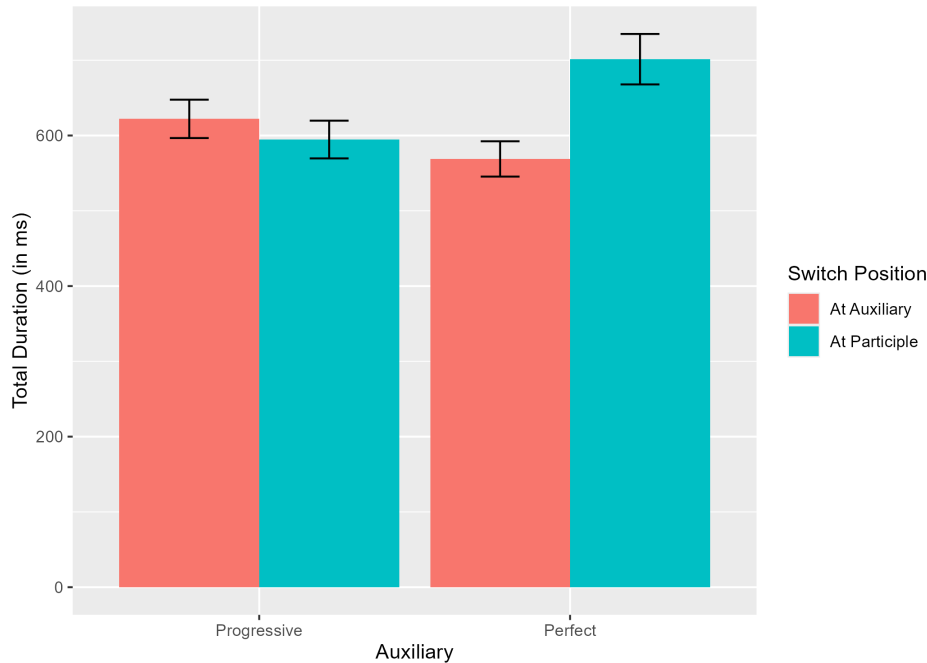


Figure 3: Mean total duration reading time in milliseconds on the spillover region (Words 9 and 10) following the critical participle region graphed by Auxiliary Type and Switch Position. Error bars represent standard error of the mean.

3.3.1 Critical word region (Word 8)

We fit a linear mixed-effects regression model for the gaze duration dataset including random intercepts for Participants. This model showed a significant main effect for Switch Position ($b = 113.62, SE = 17.96, df = 426.06, t = 6.327, p < 0.001$), a marginal effect for the Auxiliary ($b = 30.56, SE = 18.17, df = 430.16, t = 1.682, p = 0.093$), and a significant interaction between the two factors ($b = 71.47, SE = 36.3, df = 431.34, t = 1.969, p = 0.05$). The positive coefficients for the variable of Switch Position and Auxiliary indicate that participants read codeswitches that occurred at the participle more slowly as well as a trend towards reading the code-switches involving the perfect auxiliary, *haber*, more slowly. To follow-up on the interaction, we conducted pairwise tests and found several contrasting pairs that were significant. For simplification, we focus on pairs of theoretical interest from the factorial design. Comparing the switch position for each auxiliary, participants were slower to read when code-switches occurred at the participle as compared to switches at the auxiliary (for progressive

5 Sensitivity to codeswitching asymmetries in second language processing

estar trials: difference = 77.89, $SE = 20.3$, $df = 423$, $t = 3.839$, $p = 0.001$; for perfect *haber* trials: difference = 149.35, $SE = 30.2$, $df = 436$, $t = -4.944$, $p < 0.001$). Neither of the contrasts comparing between auxiliary structures but at the same code-switch position was significant (all $p > 0.133$).

The linear mixed-effects model that converged for total duration included random intercepts for Participants. The model output revealed a main effect for Switch Position ($b = 332.91$, $SE = 64.15$, $df = 488.84$, $t = 5.189$, $p < 0.001$), for Auxiliary ($b = 147.1$, $SE = 65.36$, $df = 491$, $t = 2.25$, $p = 0.025$), and a marginal interaction between the two factors ($b = 249.44$, $SE = 131.09$, $df = 491.14$, $t = 1.903$, $p = 0.058$). While the interaction was marginal, we decided to conduct pairwise tests to explore this trending interaction further. As above, there were several contrasts that were significant, but we report on only pairs that were theoretically of interest following our factorial design. Reflecting the main effect of Switch Position, participants read the critical region more slowly when code-switches occurred at the participle (for progressive *estar* trials, difference = 208.2, $SE = 69.8$, $df = 478$, $t = 2.982$, $p = 0.016$; for perfect *haber* trials, difference = 457.6, $SE = 110$, $df = 498$, $t = 4.158$, $p < 0.001$). When comparing across auxiliaries for code-switches occurring at the same syntactic juncture, only the contrast for code-switches at the participle was marginally significant, with the perfect condition being slower than the progressive condition (difference = 271.8, $SE = 112$, $df = 499$, $t = 2.428$, $p = 0.073$). Graphs for both reading measures on the critical region are provided in Figures 4 and 5.

3.3.2 Spillover region (Words 9 and 10)

In the spillover region, the linear mixed-effects regression model fit for gaze duration included a random effects structure with random intercepts on Participants and Items. This model revealed a significant effect of Switch Position ($b = 32.9$, $SE = 15.54$, $df = 314.23$, $t = 2.116$, $p = 0.035$). No effect for Auxiliary or interaction was found (Auxiliary: $b = 25.93$, $SE = 15.79$, $df = 323.77$, $t = 1.642$, $p = 0.102$; interaction: $b = 47.39$, $SE = 31.65$, $df = 312.91$, $t = 1.5$, $p = 0.135$). As in the critical region, the main effect of Switch Position indicates slower reading on trials where the code-switch occurs at the participle. Note, however, that visually, the effect seems to largely be attributable to the magnitude difference in the *haber* trials (mean difference = 74.67) rather than the difference in the *estar* trials (mean difference = 0.91), despite the lack of evidence for a significant interaction.

The model for total duration had the same random effects structure as for gaze duration. This model also only detected a significant main effect for Switch Po-

J. R. Valdés Kroff, H. G. Treadway, R. E. Guzzardo Tamargo and P. E. Dussias

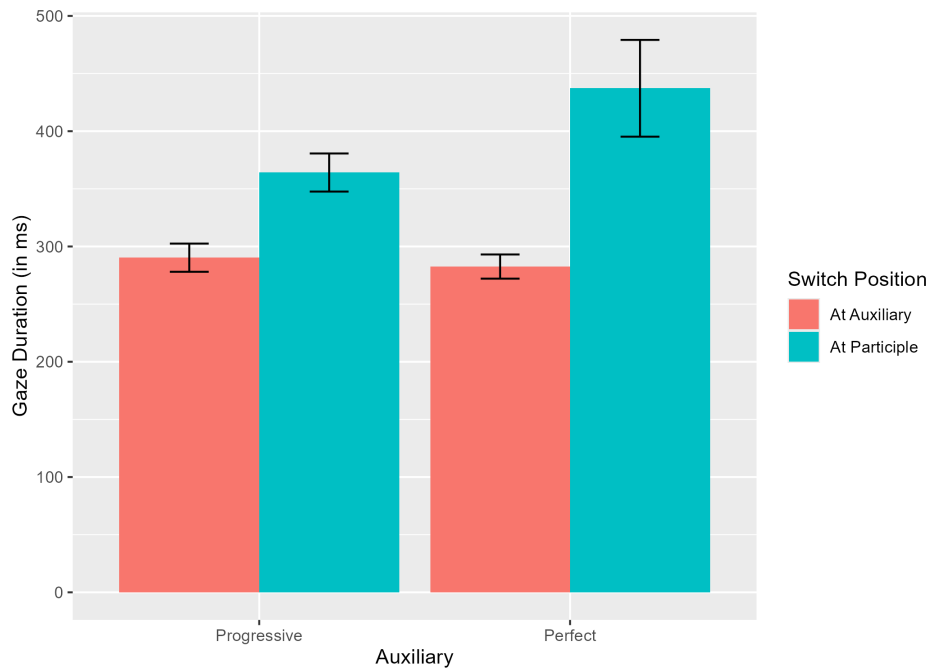


Figure 4: Mean gaze duration reading time in milliseconds on the participle critical region graphed by Auxiliary Type and Switch Position. Error bars represent standard error of the mean.

sition ($b = 129.53, SE = 50.27, df = 369.47, t = 2.576, p = 0.01$). There was no significant effect for Auxiliary or the interaction between both factors (Auxiliary: $b = 79.39, SE = 51.47, df = 379.5, t = 1.543, p = 0.124$; interaction: $b = 19.16, SE = 103.09, df = 372.42, t = 0.186, p = 0.853$). Graphs visually representing the results in the spillover region are presented in Figures 6 and 7.

In summary, a differential reading pattern emerges for the L2 Spanish group that performed a grammaticality judgment task while reading code-switched sentences. For one, this group experienced greater data loss when trimming for fixations that were not at least 150 ms. More importantly, the only consistent finding was for Switch Position where overall trials on which switches occurred at the participle were read more slowly than at the auxiliary. Potentially, a weak asymmetry was found in the early reading measure whereby for codeswitches occurring at the participle, the difference between the perfect conditions was greater than for the difference between the progressive conditions. However, a significant interaction between the variables of auxiliary and switch position was not found, indicating no clear differential sensitivity to codeswitch trials based

5 Sensitivity to codeswitching asymmetries in second language processing

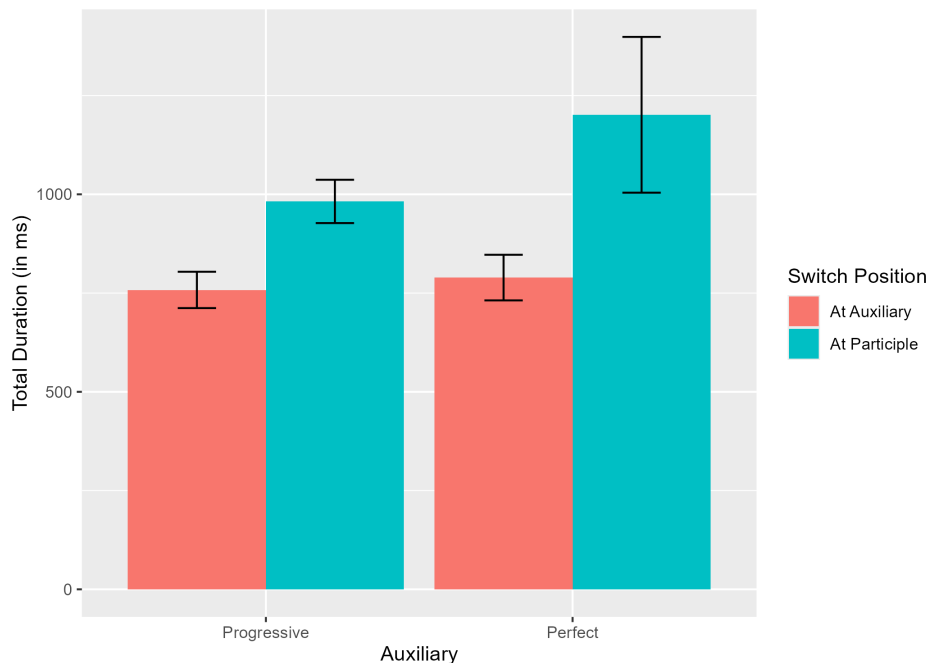


Figure 5: Mean total duration reading time in milliseconds on the participle critical region graphed by Auxiliary Type and Switch Position. Error bars represent standard error of the mean.

on the auxiliary. These results stand in contrast from the other L2 group as well as from prior reported findings.

4 Discussion and conclusions

For this study, we employed eye-tracking as a sensitive reading measure to investigate the real-time sentence processing of codeswitched sentences with complex verb phrases. Our participants were L2 Spanish participants whose L1 was English and who were immersed in an environment where the use of Spanish-English codeswitching was present. Our primary objective was to investigate whether a production asymmetry attested among Spanish-English bilinguals from the same region would emerge in this L2 Spanish group. Specifically, prior findings indicated a preference for codeswitches that occur with the progressive auxiliary, *estar*, thereby leading to frequent switching before and within the verb complex. In contrast, codeswitches with the perfect auxiliary,

J. R. Valdés Kroff, H. G. Treadway, R. E. Guzzardo Tamargo and P. E. Dussias

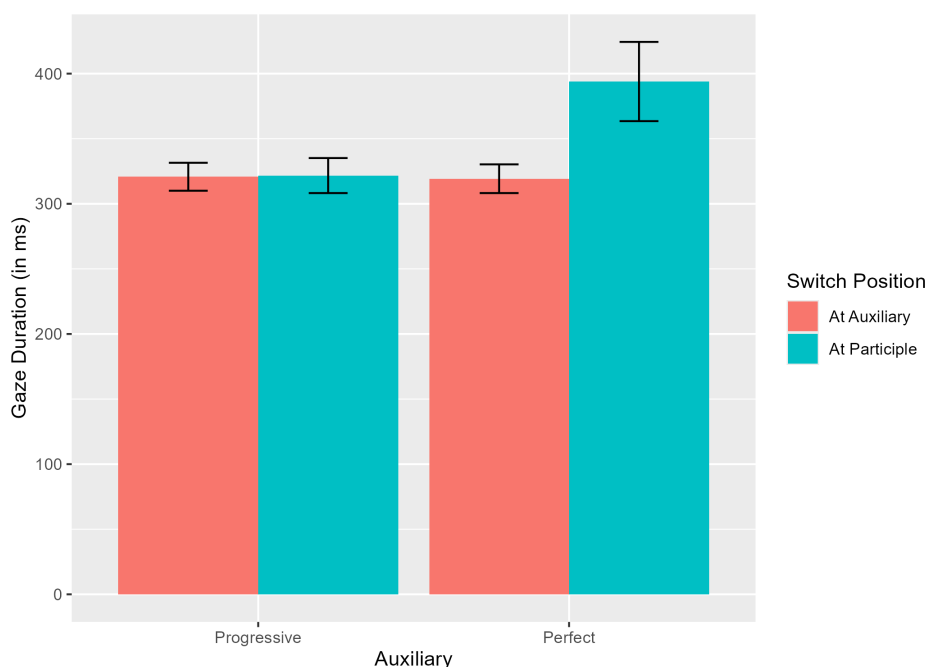


Figure 6: Mean gaze duration reading time in milliseconds on the spillover region (Words 9 and 10) graphed by Auxiliary Type and Switch Position. Error bars represent standard error of the mean.

haber, demonstrate an overwhelming preference for switching before the verb complex. This robust finding has been found in corpus studies, acceptability judgment tasks, and with eye-tracking (Guzzardo Tamargo et al. 2016).

We further investigated whether the type of behavioral task employed in experimental reading studies could influence the processing patterns of the learner group. Our results partially corroborate prior findings for Spanish-English bilinguals. Only the group that completed comprehension questions during the reading task reliably demonstrated the asymmetry in reading where perfect codeswitches that occurred at the participle were read more slowly than perfect codeswitches that occurred at the auxiliary (*los estudiantes han borrowed* v. *los estudiantes have borrowed*), a finding that was not as evident for the progressive codeswitches (*los estudiantes están borrowing* v. *los estudiantes are borrowing*). For the group that completed a forced-choice grammaticality judgment task, results pointed towards a short-lived effect of a general reading slowdown for switches at the participle. Before further unpacking the results for the comprehension question group, we discuss the different processing patterns for each L2

5 Sensitivity to codeswitching asymmetries in second language processing

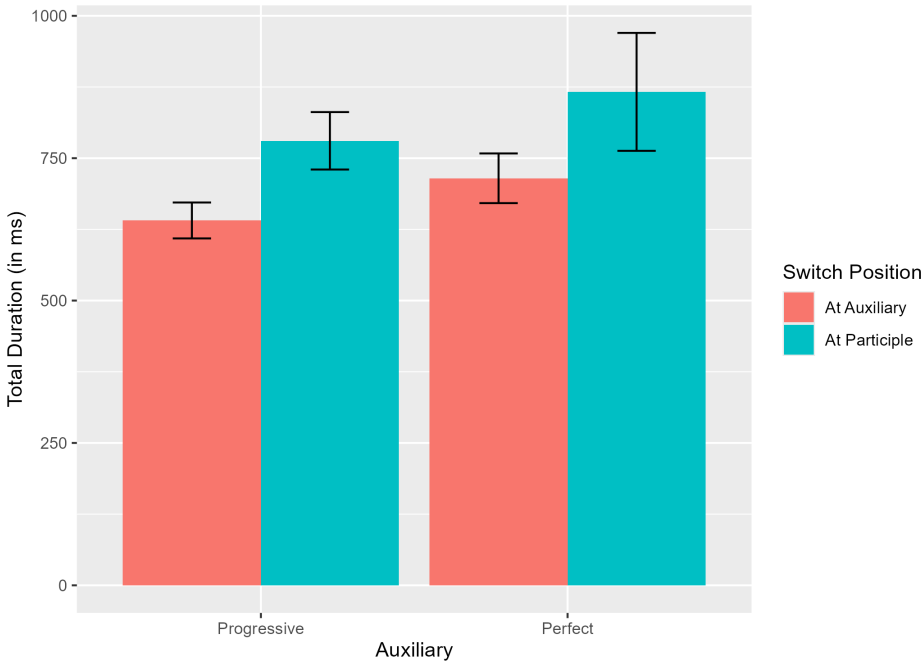


Figure 7: Mean total duration reading time in milliseconds on the spillover region (Words 9 and 10) graphed by Auxiliary Type and Switch Position. Error bars represent standard error of the mean.

group.

The results for the between-groups manipulation of secondary task are remarkably similar to prior work (Valdés Kroff et al. 2018). In the prior study, early Spanish-English bilinguals and Spanish L1-English L2 bilinguals participated in a reading study with a similar experimental design to the study reported here. For that study, secondary task was a within-group manipulation, where order of session was counterbalanced. As in the study here, the anticipated asymmetric processing pattern for perfect auxiliaries only robustly emerged for the session in which comprehension questions were employed. Instead, for the forced-choice grammaticality judgment session, participants read the sentences more slowly and no longer demonstrated an asymmetry for the perfect auxiliary verb complex. The authors attributed this processing difference to the overarching task demands for each type of task. They stated that the grammaticality judgment task induced metalinguistic mechanisms that differed from naturalistic sentence processing. Our findings thus align with the perspective that reading for comprehension engages in naturalistic processing strategies that more closely align

J. R. Valdés Kroff, H. G. Treadway, R. E. Guzzardo Tamargo and P. E. Dussias

with production distributions found in bilingual speech communities (Macizo & Bajo 2006).

Despite these similarities, we should point out that the participants from our study were different for each task session, although they came from the same population and were immersed in the same environment. The groups did not differ on a number of self-reported and assessment-based metrics in Spanish; however, the greater data loss for the grammaticality judgment task group was not expected, especially since data trimming only included removal of fixations that were less than 150 ms. Visual inspection of the eye-tracking data did not reveal any systematic differences in quality of the eye-tracking data (except for the one participant who was removed from analysis). Nevertheless, a replication with more participants in each session or using the within-participants design reported in Valdés Kroff et al. (2018) would help ensure greater confidence in the overall findings. Future studies should also consider other forms of common experimental linguistic tasks such as an acceptability judgment task with a Likert scale or visual analogue scales. Analyses that include procedures from signal detection theory such as d' could also rule out potential participant bias.

Focusing on the results for the group that completed comprehension questions during reading, we highlight two important findings in relation to prior results with early Spanish-English and L2 English bilinguals (Guzzardo Tamargo & Dussias 2013, Guzzardo Tamargo et al. 2016). First, our hypothesized asymmetry was found primarily for the late reading measure, total duration, and second, it extended into the spillover region. Prior work also robustly found the asymmetry with the early reading measure of gaze duration. This difference in timecourse suggests that the L2 participants had not yet automatized their sensitivity to the codeswitching patterns to which they are potentially exposed. Rather, their sensitivity emerges at later stages of reading that integrate regressions and re-reads of relevant regions. We can only speculate at this point if this timecourse difference is a hard constraint or if the L2 Spanish group will begin to show sensitivity in an early reading measure with increasing proficiency and/or exposure. While the effect only emerged in later stages of processing, it was a sustained effect, further underscoring that this purported asymmetry arises due to greater difficulty in integration of codeswitches within the mixed verb compound if they involve the perfect auxiliary.

Taken together, the processing differences between the L2 Spanish groups and the later and sustained emergence of the asymmetry in the timecourse of reading for participants engaged in naturalistic reading, lead us to attribute the processing of codeswitched structures to be driven more generally by experiential

5 *Sensitivity to codeswitching asymmetries in second language processing*

factors, including community norms. Specifically for the linguistic structures investigated in our study, this asymmetry arises from exposure to bilinguals who codeswitch and commonly switch in the verb phrase with the progressive auxiliary but not with the perfect auxiliary. While community preferences play an important role, the locus for the asymmetry likely derives from the grammatical constraints imposed by the participating languages, as we have argued elsewhere (Guzzardo Tamargo et al. 2016, Valdés Kroff & Dussias 2023). One such explanation may be due to the further grammaticalization of the perfect auxiliary in comparison to the progressive form in Spanish. Its status as a dependent form may very well prevent or highly disfavor codeswitches between the auxiliary and the participle.

The findings for the comprehension group subsequently offer additional evidence for a connection between production distributions and sentence processing as claimed by experience-based accounts such as the Production-Distribution-Comprehension framework (MacDonald 2013). This model claims that ease of sentence processing is directly related to the frequency of different structures in production found in a speaker community in combination with universal pressures on speech planning such as incorporating easier elements first in an utterance. Subsequently, we do not expect bilinguals from different communities to necessarily engage in the same frequency of codeswitching patterns, a claim that is supported by corpus analysis and acceptability data across different Spanish-English communities for the structures used in our study (Balam 2016, Balam et al. 2020). One prediction that stems from these comparative differences across bilingual communities is that their processing of these structures should be quite different based on community norms, independent of proficiency, a finding that we have begun to explore (Treadway et al. 2025). A comparative sentence processing approach will be necessary to offer key evidence to test our hypothesis. If other bilingual communities such as Spanish-English bilinguals in Belize who demonstrably show frequent use of different verb phrase code-switches nevertheless show the asymmetric sensitivity reported here, then the role of community norms and exposure would not be as important as we claim.

One prediction that stems from these comparative differences across bilingual communities is that their processing of these structures should be quite different based on community norms—independent of proficiency—a finding that we have begun to explore (Treadway et al. 2025). A comparative sentence processing approach will be necessary to offer key evidence to test our hypothesis. If other bilingual communities such as Spanish-English bilinguals in Belize who demonstrably show frequent use of different verb phrase code-switches nevertheless show the asymmetric sensitivity reported here, then the role of community

J. R. Valdés Kroff, H. G. Treadway, R. E. Guzzardo Tamargo and P. E. Dussias

norms and exposure would not be as important as we claim. This comparative approach fundamentally depends on the availability of bilingual corpora from different communities.

Acknowledgements

The writing of this article was supported by the U.S. National Science Foundation (NSF) while the first author worked at the Foundation. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. This project was partly supported by NSF OISE 1545900 and NSF NRT 2125865 to Paola E. Dussias.

References

- Aaronson, Doris & Steven Ferres. 1986. Reading strategies for children and adults: A quantitative model. *Psychological Review* 93(1). 89–112. DOI: [10.1037/0033-295X.93.1.89](https://doi.org/10.1037/0033-295X.93.1.89).
- Avery, Nick & Emma Marsden. 2019. A meta-analysis of sensitivity to grammatical information during self-paced reading: Towards a framework of reference for reading time effect sizes. *Studies in Second Language Acquisition* 41(5). 1055–1087. DOI: [10.1017/S0272263119000196](https://doi.org/10.1017/S0272263119000196).
- Balam, Osmer. 2016. Mixed verbs in contact Spanish: Patterns of use among emergent and dynamic bi/multilinguals. *Languages* 1(1). 3. DOI: [10.3390 / languages1010003](https://doi.org/10.3390/languages1010003).
- Balam, Osmer, Maria del Carmen Parafita Couto & Hans Stadthagen-González. 2020. Bilingual verbs in three Spanish/English code-switching communities. *International Journal of Bilingualism* 24(5–6). 952–967. DOI: [10.1177 / 1367006920911449](https://doi.org/10.1177 / 1367006920911449).
- Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68. 255–278. DOI: [10.1016/j.jml.2012.11.001](https://doi.org/10.1016/j.jml.2012.11.001).
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48. DOI: [10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01).
- Beatty-Martínez, Anne L., Jorge R. Valdés Kroff & Paola E. Dussias. 2018. From the field to the lab: A converging methods approach to the study of codeswitching. *Languages* 3(2). 952–967.

5 Sensitivity to codeswitching asymmetries in second language processing

- Belazi, Hedi M., Edward J. Rubin & Almeida Jacqueline Toribio. 1994. Code switching and x-bar theory: The Functional Head Constraint. *Linguistic Inquiry* 25. 221–237. <https://www.jstor.org/stable/4178859>.
- Carter, Phillip M. & Aandrew Lynch. 2015. Multilingual Miami: Current trends in sociolinguistic research. *Language and Linguistics Compass* 9(9). 369–385. DOI: [10.1111/lnc3.12157](https://doi.org/10.1111/lnc3.12157).
- Dell, Gary S. & Franklin Chang. 2014. The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1634). 20120394. DOI: [10.1098/rstb.2012.0394](https://doi.org/10.1098/rstb.2012.0394).
- Deuchar, Margaret. 2020. Code-switching in linguistics: A position paper. *Languages* 5(2). 22. DOI: [10.3390/languages5020022](https://doi.org/10.3390/languages5020022).
- Deuchar, Margaret, Peredur Davies, Jon Russell Herring, Maria del Carmen Parafta Couto & Diana Carter. 2014. Building bilingual corpora. In Enlli Môn Thomas & Ineke Mennen (eds.), *Advances in the study of bilingualism*, 93–110. Bristol: Multilingual Matters. DOI: [10.21832/9781783091713-008](https://doi.org/10.21832/9781783091713-008).
- Di Sciullo, Anne-Marie, Pieter Muysken & Rajendra Singh. 1986. Government and code-mixing. *Journal of Linguistics* 22(1). 1–24. DOI: [10.1017/S0022226700010537](https://doi.org/10.1017/S0022226700010537).
- Dussias, Paola E. 2003. Spanish-English code mixing at the Auxiliary Phrase: Evidence from eye-movement data. *Revista Internacional de Lingüística Iberoamericana* 1(2). 7–34. <https://www.jstor.org/stable/41678169>.
- Dussias, Paola E., Jorge R. Valdés Kroff, Rosa E. Guzzardo Tamargo & Chip Gergen. 2013. When gender and looking go hand in hand: Grammatical gender processing L2 Spanish. *Studies in Second Language Acquisition* 35. 353–387. DOI: [10.1017/S0272263112000915](https://doi.org/10.1017/S0272263112000915).
- Fodor, Jerry A. & Thomas G. Bever. 1965. The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior* 4(5). 414–420. DOI: [10.1016/S0022-5371\(65\)80081-0](https://doi.org/10.1016/S0022-5371(65)80081-0).
- Giancaspro, David. 2015. Code-switching at the auxiliary-VP boundary: A comparison of heritage speakers and L2 learners. *Linguistic Approaches to Bilingualism* 5(3). 379–407. DOI: [10.1075/lab.5.3.04gia](https://doi.org/10.1075/lab.5.3.04gia).
- Godfroid, Aline. 2019. *Eye-tracking in second language acquisition and bilingualism*. New York: Taylor & Francis. DOI: [10.4324/9781315775616](https://doi.org/10.4324/9781315775616).
- Gollan, Tamar H. & Matthew Goldrick. 2016. Grammatical constraints on language switching: Language control is not just executive control. *Journal of Memory and Language* 90. 177–199. DOI: [10.1016/j.jml.2016.04.002](https://doi.org/10.1016/j.jml.2016.04.002).

J. R. Valdés Kroff, H. G. Treadway, R. E. Guzzardo Tamargo and P. E. Dussias

- Guzzardo Tamargo, Rosa E. & Paola E. Dussias. 2013. Processing of Spanish-English code-switches by late bilinguals. In Sarah Baiz, Nora Goldman & Rachel Hawkes (eds.), *BUCLD 37: Proceedings of the 37th Annual Boston University Conference on Language Development*, vol. 37, 134–146. Somerville: Cascadilla Press.
- Guzzardo Tamargo, Rosa E., Jorge R. Valdés Kroff & Paola E. Dussias. 2016. Examining the relationship between comprehension and production processes in code-switched language. *Journal of Memory and Language* 89. 138–161. DOI: [10.1016/j.jml.2015.12.002](https://doi.org/10.1016/j.jml.2015.12.002).
- Joshi, Aravind K. 1985. Processing of sentences with intrasentential code switching. In David R. Dowty, Lauri Karttunen & Arnold M. Zwicky (eds.), *Natural language parsing: Psychological, computational and theoretical perspectives*, 190–205. Cambridge: Cambridge University Press. DOI: [10 . 1017 / CBO9780511597855.006](https://doi.org/10.1017/CBO9780511597855.006).
- Kaakinen, Johanna K. & Jukka Hyönä. 2010. Task effects on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36(6). 1561–1566. DOI: [10.1037/a0020693](https://doi.org/10.1037/a0020693).
- Koronkiewicz, Bryan. 2018. Acquiring L1-English L2-Spanish code-switching: The role of exposure to language mixing. *Languages* 3(3). 26. DOI: [10.3390 / languages3030026](https://doi.org/10.3390/languages3030026).
- Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26. DOI: [10.18637/jss.v082.i13](https://doi.org/10.18637/jss.v082.i13).
- Lenth, Russell V. 2024. *emmeans: Estimated marginal means, aka least-squares means*. R package version 1.10.2. <https://CRAN.R-project.org/package=emmeans>.
- Liceras, Juana M., Raquel Fernández Fuertes, Susana Perales, Rocío Pérez-Tattam & K. Todd Spradlin. 2008. Gender and number agreement in bilingual native and non-native grammars: A view from child and adult functional-lexical mixings. *Lingua* 118(6). 761–852. DOI: [10.1016/j.lingua.2007.05.006](https://doi.org/10.1016/j.lingua.2007.05.006).
- Lipski, John M. 1978. Code switching and the problem of bilingual competence. In Michel Paradis (ed.), *Aspects of bilingualism*, 250–264. Columbia: Hornbeam Press.
- Lipski, John M. 1985. *Linguistic aspects of Spanish-English language switching*. Tucson: Center for Latin American Studies, Arizona State University.
- Lohndal, Terje & Michael T. Putnam. 2024. The importance of features and exponents: Dissolving feature reassembly. *Linguistic Approaches to Bilingualism* 14(1). 1–36.

5 Sensitivity to codeswitching asymmetries in second language processing

- López, Luis. 2020. *Bilingual grammar: Toward an integrated model*. Cambridge: Cambridge University Press.
- López, Luis, Artemis Alexiadou & Tonjes Veenstra. 2017. Code-switching by phase. *Languages* 2(3). 9. DOI: [10.3390/languages2030009](https://doi.org/10.3390/languages2030009).
- MacDonald, Maryellen C. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology* 4. 226. DOI: [10.3389/fpsyg.2013.00226](https://doi.org/10.3389/fpsyg.2013.00226).
- Macizo, Pedro & M. Teresa Bajo. 2006. Reading for repetition and reading for translation: Do they involve the same processes? *Cognition* 99(1). 1–34. DOI: [10.1016/j.cognition.2004.09.012](https://doi.org/10.1016/j.cognition.2004.09.012).
- MacSwan, Jeff. 1999. *A minimalist approach to intrasentential code switching*. London: Routledge. DOI: [10.4324/9781315053158](https://doi.org/10.4324/9781315053158).
- MacSwan, Jeff. 2005. Codeswitching and generative grammar: A critique of the MLF model and some remarks on “modified minimalism”. *Bilingualism: Language and Cognition* 8(1). 1–22. DOI: [10.1017/S1366728904002068](https://doi.org/10.1017/S1366728904002068).
- MacSwan, Jeff. 2014. Programs and proposals in codeswitching research: Unconstraining theories of bilingual language mixing. In Jeff MacSwan (ed.), *Grammatical theory and bilingual codeswitching*, 1–33. Cambridge: MIT Press.
- Mahootian, Shahrzad. 1996. Code-switching and universal constraints: Evidence from Farsi/English. *World Englishes* 15(3). 377–384. DOI: [10.1111/j.1467-971X.1996.tb00124.x](https://doi.org/10.1111/j.1467-971X.1996.tb00124.x).
- Miccio, Adele W., Carol Scheffner Hammer & Bárbara Rodríguez. 2009. Code-switching and language disorders in bilingual children. In Barbara E. Bullock & Almeida Jacqueline Toribio (eds.), *The Cambridge handbook of linguistic code-switching*, 241–252. Cambridge: Cambridge University Press. DOI: [10.1017/CBO9780511576331.015](https://doi.org/10.1017/CBO9780511576331.015).
- Myers-Scotton, Carol. 1993. *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon.
- Myers-Scotton, Carol & Janice L. Jake. 2000. Testing the 4-M model: An introduction. *International Journal of Bilingualism* 4(1). 1–8. DOI: [10.1177/1367006900040010101](https://doi.org/10.1177/1367006900040010101).
- Myers-Scotton, Carol & Janice L. Jake. 2001. Explaining aspects of code-switching and their implications. In Janet L. Nicol (ed.), *One mind, two languages: Bilingual language processing*, 84–116. London: Blackwell.
- Myers-Scotton, Carol & Janice L. Jake. 2017. Revisiting the 4-M model: Codeswitching and morpheme election at the abstract level. *International Journal of Bilingualism* 21(3). 340–366. DOI: [10.1177/1367006915626588](https://doi.org/10.1177/1367006915626588).

J. R. Valdés Kroff, H. G. Treadway, R. E. Guzzardo Tamargo and P. E. Dussias

- Parafita Couto, Maria del Carmen, Kate Bellamy & Felix K. Ameka. 2023. Theoretical linguistic approaches to multilingual code-switching. In Jennifer Cabrelli, Adel Chaouch-Orozco, Jorge González Alonso, Sergio Miguel Pereira Soares, Eloi Puig-Mayenco & Rothman Jason (eds.), *The Cambridge handbook of third language acquisition* (Cambridge Handbooks in Language and Linguistics), 403–436. Cambridge: Cambridge University Press. DOI: [10.1017/9781108957823.017](https://doi.org/10.1017/9781108957823.017).
- Parafita Couto, Maria del Carmen & Marianne Gullberg. 2019. Code-switching within the noun phrase: Evidence from three corpora. *International Journal of Bilingualism* 23(2). 695–714. DOI: [10.1177/1367006917729543](https://doi.org/10.1177/1367006917729543).
- Pfaff, Carol. 1979. Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language* 55. 291–318. DOI: [10.2307/412586](https://doi.org/10.2307/412586).
- Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics* 18(7/8). 581–618. DOI: [10.1515/ling.1980.18.7-8.581](https://doi.org/10.1515/ling.1980.18.7-8.581).
- Poplack, Shana. 1987. Contrasting patterns of code-switching in two communities. In Erling Wande, Jan Anward, Bengt Nordberg, Lars Streensland & Mats Thelander (eds.), *Aspects of multilingualism: Proceedings from the Fourth Nordic Symposium of Bilingualism* (Acta Universitatis Upsaliensis), 51–77. Uppsala: Uppsala Universitet.
- Rayner, Keith. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124(3). 372–422. DOI: [10.1037/0033-2909.124.3.372](https://doi.org/10.1037/0033-2909.124.3.372).
- Toribio, Almeida Jacqueline. 2001. On the emergence of bilingual code-switching competence. *Bilingualism: Language and Cognition* 4. 203–231.
- Torres Cacoullous, Rena & Catherine E. Travis. 2018. *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge: Cambridge University Press.
- Treadway, Hannah G., Josh L. Higdon & Jorge R. Valdés Kroff. 2025. Disentangling grammar and experience: On the role of environmental exposure to Spanish-English code-switching. *Isogloss* 11(4). Bilingualism in the Hispanic and Lusophone World (BHL) 4: Theoretical and experimental approaches. DOI: [10.5565/rev/isogloss.519](https://doi.org/10.5565/rev/isogloss.519).
- Valdés Kroff, Jorge R. & Paola E. Dussias. 2023. Production, processing, and prediction in bilingual codeswitching. In Kara D. Federmeier & Jessica L. Montag (eds.), *Speaking, writing and communicating* (Psychology of learning and motivation 78), 195–237. Cambridge: Elsevier. DOI: [10.1016/bs.plm.2023.02.004](https://doi.org/10.1016/bs.plm.2023.02.004).

5 Sensitivity to codeswitching asymmetries in second language processing

- Valdés Kroff, Jorge R., Paola E. Dussias, Chip Gerfen, Lauren Perrotti & M. Teresa Bajo. 2017. Experience with code-switching modulates the use of grammatical gender during sentence processing. *Linguistic Approaches to Bilingualism* 7(2). 163–198. DOI: [10.1075/lab.15010.val](https://doi.org/10.1075/lab.15010.val).
- Valdés Kroff, Jorge R., Rosa E. Guzzardo Tamargo & Paola E. Dussias. 2018. Experimental contributions of eye-tracking to the understanding of comprehension processes while hearing and reading code-switches. *Linguistic Approaches to Bilingualism* 8. 98–133. DOI: [10.1075/lab.16011.val](https://doi.org/10.1075/lab.16011.val).
- Valdés Kroff, Jorge R., Souad Kheder, Aleksandra Tomić & Edith Kaan. 2023. Psycholinguistics of multilingual code-switching. In Jennifer Cabrelli, Adel Chaouch-Orozco, Jorge González Alonso, Sergio Miguel Pereira Soares, Eloi Puig-Mayenco & Jason Rothman (eds.), *The Cambridge handbook of third language acquisition and processing*, 437–465. Cambridge: Cambridge University Press. DOI: [10.1017/9781108957823.018](https://doi.org/10.1017/9781108957823.018).
- Van Hell, Janet G. 2022. Code-switching. In Aline Godfroid & Holger Hopp (eds.), *Routledge handbook of second language acquisition and psycholinguistics*, 255–267. London: Routledge. DOI: [10.4324/9781003018872-24](https://doi.org/10.4324/9781003018872-24).

Chapter 6

Linguistic interaction in early bilingual development: A review of studies examining single- and mixed-language utterances

Marilyn M. Vihman & Virve-Anneli Vihman

Over the past 25 years or so a steady stream of papers has addressed bilingual development and the possible effects – or lack of effects – of cross-linguistically differing structures on a bilingual child's acquisition of those structures. Drawing generalizations has proven difficult, partly due to the many variables that enter in. Furthermore, the theoretical assumptions of the authors vary from one study to the next; this can make it difficult to evaluate the different conclusions drawn. A few studies have specifically focused on the long-standing puzzle of how best to weigh the effects of (i) experience with adult input patterns and (ii) the child's own relative proficiency in each language on the diverse manifestations of bilingual child usage, including code-switching. Here we provide an overview of the findings of the bilingual child literature on the extent of linguistic interaction in the acquisition of morphosyntactic structure. We review 47 studies of the spontaneous language use of bilingual children up to age 4 and emphasise the point that child production is a moving target that necessarily reflects the dynamics of the child's own emergent grammatical systems and possible shifts in their exposure to and use of the languages in question.

1 Introduction

A central concern of studies of bilingual language development is the nature and extent of interaction – also termed interference (e.g., [Paradis & Navarro 2003](#)),

Marilyn M. Vihman & Virve-Anneli Vihman

transfer (e.g., Yip & Matthews 2000) or influence (e.g., Serratrice 2013) – between the linguistic structures of the two languages to which the child has been exposed. This is crucial for understanding the process of bilingual acquisition, in which the child’s neural, cognitive and linguistic systems develop in interaction with two grammatical systems. Clarifying how cross-linguistic interaction operates in bilingual children is also important for better understanding adult bilingual processing, and for adapting models of bilingual language use.

Evidence of interaction can most readily be drawn from the child’s speech production; such interaction indicates an effect of bilingual experience on the child’s developing grammatical systems. Here we provide an overview of studies from the last three decades that address cross-linguistic interaction in bilingual children’s production during the period of emergence of morphosyntax. We understand code-switching (or the production of ‘mixed-language utterances’) to constitute one manifestation of such interaction. Note, moreover, that the evidence we investigate from mixed utterances is not the inclusion of elements from more than one language *per se*, but rather the effect of the structure of one language on production in the other, the same kind of evidence we investigate in single-language utterances. As cross-linguistic influence may manifest differently in utterances with lexemes from one language compared to mixed-language utterances, we divide our sample of studies into (i) those that set mixed utterances aside to focus on cross-linguistic influence in single-language utterances and (ii) those that specifically address utterances that include both languages.

Many factors conspire to make the effects of linguistic interaction on bilingual language use a challenging research topic. In adults such factors include the difficulty of measuring and controlling for heterogeneity in proficiency, usage and relative dominance in the two languages, differences in their typological relatedness and also task differences that might affect executive control, cross-language cues and language activation (see Grosjean 1998, Baum & Titone 2014). For developmental bilingualism, a critical issue is the *extent of exposure*, or *opportunities for hearing and using each language* – and this is liable to *change over time* (cf. Tulloch & Hoff 2023). *Language dominance* – the child’s relative proficiency in each language – is affected by relative exposure to and opportunities for use of the two languages but also interacts with the child’s *overall cognitive and linguistic development*. In young children, furthermore, it is important to consider the wide range of *individual differences*, which are also commonly observed in monolingual development (cf. Serratrice 2005, who explores the effect of individual differences on learning strategy in two bilingual children raised in similar circumstances, though with different language pairs).

6 Linguistic interaction in early bilingual development

Beyond these sources of complexity for analysis, the differing *theoretical assumptions* of authors also make it difficult to draw generalizable conclusions from disparate studies. We note further that *language usage in the community* is seldom taken into account, with the exception of a few recent studies (e.g., Yip & Matthews 2016: Cantonese & English in Hong Kong; Qi & Di Biase 2020: Mandarin and English in Australia; Balam et al. 2021 and van Osch et al. 2023: Spanish and English in Miami, Florida and the Netherlands, resp.; Phillips & Deuchar 2022: Welsh and English in North Wales). The linguistic context can be taken to include family language policy, community norms, the extent of bilingual usage in the community, community attitudes toward bilingualism and/or toward each language and the pattern of interaction of the two languages in the community. For example, Yip & Matthews (2016) mention that although code-switching is widespread in Hong Kong, families there vary in extent of switching, and this influences the children's usage. Phillips & Deuchar (2022) specifically investigate how the child's pattern of switching is related to the prevailing adult usage in a bilingual Welsh-English community, which resembles Hong Kong in the prevalence of switching. The linguistic context, in both the home and the larger community, is relevant for the question of cross-language influence as well as for understanding developmental code-switching patterns.

2 Background

Over the past 30-odd years the number of investigators interested in documenting and discussing the process of learning two languages in early childhood has expanded considerably, and the study of bilingual language development has surged. Our initial search on Google Scholar (February 2023), using the simple term 'bilingual language development' and limited to studies from 1990 on, yielded nearly 5000 titles. Kroll & Bialystok (2013, p. 498) document a 'steep and continuing rise in both publications and citations' on bilingualism in general from 2003 on; we see a similar increase in bilingual development, with half of the responses from our initial search dating to 2010 or later. To address the question of cross-linguistic interaction in bilingual children in a systematic, quantitative manner, Van Dijk et al. (2022) recently provided a meta-analysis of experimental studies, restricted to those with explicit measurement and at least two bilingual and two monolingual children. All participants were aged three years or older. That paper examines evidence for cross-linguistic influence from 26 studies, with data from 750 children and 17 unique language pairs (Hebrew is the only non-European language included). These studies typically involve elicited

Marilyn M. Vihman & Virve-Anneli Vihman

production, with a few testing grammaticality judgments and a small number involving comprehension only. Four-year-olds were the group most often tested, but school-age children up to the age of ten were also included.

The Van Dijk et al. study is systematic and thorough, but the focus on older children and experimental results leaves a significant gap. Younger children may be more variable and difficult to include in well-controlled experiments but assessing the extent of separation or cross-linguistic interaction in their emergent grammars is essential for forming an adequate picture of the trajectory of bilingual grammatical development. For very young children, spontaneous production provides the most reliable information, reflecting their behavior in a natural, everyday context. Here, then, we provide a complementary overview of studies investigating data (i) from the early years of morphosyntactic development, (ii) in spontaneous speech production. Considering children's early use of their developing languages should enable us to detect the effect of the relative accessibility of structures, their parallel development and the extent of bilingual interaction as the two grammatical systems first emerge and gradually come into regular use. To achieve these ends we focus on studies that specify cross-linguistic effects on particular morphosyntactic structures, in either single- or mixed-language utterances.

In a second search, following Van Dijk et al. (2022), we entered 'bilingual child' as our basic search term, combined with {'cross-language'} or {'cross-linguistic'} and {'effects', 'transfer', 'influence', 'interaction', 'activation'}. Together these search terms yielded over 9000 titles, with a certain amount of overlap. We then refined the results by restricting our sample to studies providing *data from spontaneous speech* (based on transcripts of audio- or video-recordings, diary notes, or both) and relating to *morphosyntax*. In addition, for comparability, we chose sources that provide data derived from (i) *typically developing children*, (ii) recorded from *the age of four or younger*, (iii) exposed to *two oral languages* and (iv) including *morphosyntactic analysis of multiword utterances*, whether those include code-switching or not. We thus excluded studies using elicited data or focusing solely or mainly on phonetics, phonology, the lexicon, sign languages or bimodal bilingualism, as well as studies deriving data from children with atypical development; we also excluded overviews, whether purely theoretical or oriented toward teaching or clinical use. Finally, we included only studies that provide *explicit analysis and discussion of morphosyntactic interaction*, in single- or mixed-language utterances, and with specific reference to the morphosyntactic structures affected. With those restrictions, and with the addition of further references derived from the readings, we identified 47 empirical studies. Note, further,

6 *Linguistic interaction in early bilingual development*

that our search was restricted to articles published in English, admittedly a limitation for a study seeking information on typologically varied language pairs.

Linguistically, the studies are wide-ranging: Overall, 21 different languages and 30 unique language pairs are included in at least one study each. English is one of the languages of a pair in 34 studies, German in 23, French or Italian in 15 each, Spanish in 10; note that some studies report and analyze data from more than one language pair and also that some of the same children's data are treated in more than one study, with a focus on different structures. The 47 studies include 197 children altogether, or a mean of 4.2 children per study. There are 30 single-child case studies; 14 studies include at least four children learning one or more of the language pairs they address; only three include more than ten. In short, these studies, selected for their analyses of early spontaneous bilingual language use, generally report data from very few children.¹ The review is largely qualitative, given both the type of data available for younger children and the limited numbers of participants per study.

At least three key issues in the literature on bilingual language development have sharply changed over the last three decades. First, attitudes and assumptions have changed: In 1988 Petersen still felt it necessary to insist that her daughter's addition of English inflections to Danish content nouns did not reflect '*random lexical choices*' or '*confusion*' (as the popular view might hold: Goodz 1989), but instead could be seen to reveal *systematic patterning*. Denials that bilingual children are 'confused' have continued to surface from time to time (e.g., Yip & Matthews 2000: 207; Paradis & Navarro 2003: 372), but there has long been general agreement on the systematicity of child code-switching, in accord with the consensus, securely established over this same period, that adult code-switching is itself systematic (cf., e.g., Kroll & De Groot 2005, Deuchar 2013, 2020, Valdés Kroff & Dussias 2023).

Secondly, in the late 1980s the question arose as to whether the bilingual child *learns from the input* to use intraclausal code-switching, or code-mixing (the '*modeling hypothesis*': Comeau et al. 2003). Goodz (1989) reported that, in a longitudinal study of four families in Montreal, the input the child heard tended to include switching, despite the parents' stated commitment to each of them using their native language only. The proportion of mixed language utterances was low, however, and tended to be initiated by the child and then echoed by the

¹Sağın-Şimşek & Antonova-Ünlü (2021) is the only study included here to have used elicitation – namely, of the Frog Story narrative (Mayer 1969) – as part of their (annual) data collection; this was supplemented with recordings of naturalistic play. The children – four monolinguals and four in each of three bilingual groups (English-, German- and Russian-Turkish) – ranged in age from 2;11 to 3;11 in the first recordings, at the upper limit of our sample.

Marilyn M. Vihman & Virve-Anneli Vihman

parent rather than the reverse. In a study designed to test the question, [Genesee et al. \(1995\)](#) found no evidence that the language mixing of the five bilingual children they recorded with each parent was correlated with the extent of mixing in their parents' speech. Mixing in parental input now seldom seems to be invoked as a source of child code-switching (see [Meng & Miyamoto 2012](#), [Tulloch & Hoff 2023](#)).

A closely related aspect of the issue of learning from the input is the question of parental strategy. Over half of the studies we review here (27 of 46) report that the participating families observe the 'one person-one language' principle in raising a bilingual child (as did the families recorded in the studies of both Goodz and Genesee et al. in Montreal). The principle was originally recommended by [Ronjat \(1913\)](#); it was later forcefully promoted by investigators such as [Döpke \(1992\)](#) and [De Houwer \(1995, 2005\)](#). Yet in the papers we review parental strategy is seldom mentioned as a factor contributing to either linguistic interaction or the lack of it. One exception is [Yip & Matthews \(2016\)](#), who report a higher level of mixing in the production of children from families freely using both Cantonese and English in the home, with considerable adult mixing, as compared to those raised in homes following the 'one person, one language' principle, which is less common in present-day Hong Kong. It is worth noting that where the home language differs from that of a uniformly majority-language community, many families consider primary use of the minority language in the home to be essential to ensure bilingual development (e.g., [Deuchar & Quay 2000](#), [Vihman 1985, 1998](#); cf. also [Gawlitczek-Maiwald & Tracy 1996](#): 908).

In general, the pattern of *language use in the community* should be taken into account ([Phillips & Deuchar 2022](#)). [Byers-Heinlein \(2013\)](#), a large-scale study of parental input and its effect on vocabulary growth, reports considerable mixing in bilingual families in Vancouver, most of whom are part of a large bilingual community. Beyond that, [Phillips & Deuchar \(2022\)](#) show that in North Wales, where the minority language (Welsh) consistently provides the matrix into which content words from the majority language (English) are inserted in the commonly occurring code-switching of adults, the children's mixed utterances follow suit; the reverse pattern – with English as the language of the verb and function words, Welsh as the source of content-word insertions – is virtually unattested in the corpus. Attention to the language of the community might also help to account for [Petersen's \(1988\)](#) findings (cf. also [Lanza 1997b,a](#), [Vihman 1998](#), for example). Where the language of the home differs from that of the community, the 'matrix language' – the language in which the predicate and most of the function words are expressed ([Myers-Scotton 1993](#)) – is most likely to be that of the home, while

6 Linguistic interaction in early bilingual development

the community language serves as a source of content-word insertions, at least in the preschool years.

Finally, the assertion that children are *able to differentiate between their two grammars* from the beginnings of word combination, if not sooner, runs as a thread through much of the literature. Genesee (1989) appealed to the considerable evidence of infants' precocious capacity for distinguishing between speech sounds to shed doubt on the claim that bilingual children go through a stage of treating their two languages as if they constituted a single unified or 'fused' system (Volterra & Taeschner 1978; see Vihman 2014, Ch. 3, for a review of the extensive infant speech perception literature supporting Genesee's point, and Byers-Heinlein 2014, for a reassessment of the issue of early language differentiation). More recent perception studies have provided ample experimental evidence that infants are able to discriminate between languages even at birth (see Nazzi et al. 1998, Ramus 2002, Byers-Heinlein et al. 2010, and that bilingual children can segment speech in both their languages by the middle of the first year, given sufficient exposure (e.g., Bosch et al. 2014, Polka et al. 2017). Accordingly, children must be sensitive from their earliest months to the presence of more than one language in their environment.

The claim that the child has 'two systems from the start' is also rooted in the theoretical clash that has pervaded acquisition studies since the publication of Chomsky's *Aspects of the Theory of Syntax* (1965), with its proposal of an innate Language Acquisition Device. If, as generative approaches have traditionally maintained, children set the parameters appropriate for each language by accessing Universal Grammar, then the two linguistic systems should develop autonomously; according to this view, there is no reason to expect interaction, at least not at the level of 'competence', or the child's underlying knowledge (see Paradis & Genesee 1996, who defend this position).

The claim of independence for each of a child's grammatical systems continues to be reiterated in the literature, even though no one has defended the 'fused system' hypothesis for over 30 years. As an example of current views, Kupisch (2007) provides evidence of interaction in the acquisition of Italian and German determiners that goes against the assumption that bilingual children have 'two autonomous grammatical systems', but she maintains that language influence (or interaction) and language separation are not mutually exclusive (p. 76). Most investigators today would probably agree with this more nuanced conclusion.

An alternative perspective on the issue, from a usage-based theoretical position, is provided by Gaskins et al. (2021). Each of a child's experiences with hearing and producing language leaves its memory trace, or set of exemplars, in

Marilyn M. Vihman & Virve-Anneli Vihman

the child's mind and thus adds to the child's stock of linguistic knowledge. Previously experienced lexical or phrasal units provide potential material for linguistic expression. By this account, both frequency-based entrenchment and 'priming', or the experience of the immediately preceding speech, may explain a child's recourse to units from one language or the other or from both within the same utterance. Gaskins et al. (2022) offer some suggestions for future research to test the validity of these ideas. In effect, most researchers might now agree that the systems emerge and develop in parallel in a child with sufficient exposure to and opportunities for use of two languages; thus some degree of linguistic interaction can be expected.

One way to understand how linguistic systems may interact in the child is the suggestion that the acquisition of a more easily learned construction in one language supports learning of the equivalent structure in the other. The proposal of linguistic facilitation, which Bernardini & Schlyter (2004) termed the 'ivy hypothesis', derives from evidence that bilingual children sometimes benefit from *accelerated learning* of aspects of grammar that are comparable in their two languages (e.g., Kupisch 2007). In contrast, the identification of difficulty with a construction as compared (by proportion of recorded attempts) with monolingual children can suggest some *delay* due to the child's experience of two systems. The errors in monolingual children's production can provide a useful developmental comparison point for bilingual children's production. However, in cases where this involves comparing a single bilingual and a single monolingual child, the findings are not generalizable, given the inherent variability in age and trajectory of individual children's linguistic advance.

Rothman et al. (2023) problematize the use of monolingual children as controls in studies of bilingual development, observing that, for example, a monolingual 'homeland' speaker group is not necessarily a revealing comparison for heritage language speakers. They note, however, that 'if one is interested in documenting the (potential) role that crosslinguistic influence has in the development of bilingual grammars in childhood, it could be reasonable to compare a child bilingual group to a monolingual one' (2023: 319), as long as it is motivated by the research question. The small sample size in the studies discussed here means that any comparisons are indicative rather than conclusive. Furthermore, where a comparison indicates delay, this may derive as much from limited input or exposure as from cross-linguistic interaction. For identifying interaction in small groups of participants, it might be more fruitful to compare bilingual groups with differing language pairs.

Another type of evidence of interaction derives from a bilingual child's mixed-language utterances, which often show reliance on function words or inflections

6 Linguistic interaction in early bilingual development

from one language along with content words from the other; such usage suggests memory or representational challenges. Note that the mixing of content words from the larger community need not mean that the child has not yet learned the relevant words in the home language or, in other words, that there is an actual ‘lexical gap’; instead, such mixing may result from a momentary ‘retrieval problem’ (Lanvers 2001: 449) or ‘retrieval failure’ (Serratrice 2005: 169). Certainly momentary lapses in access are not unknown in adult speakers, whether bilingual or not. Such lapses likely account for much of the variability in child production at every level.

We focus here on morphosyntactic interaction in early bilingual child production, dividing our overview into the two major strands we find in the literature: (i) evidence of interaction in single-language utterances in either language and (ii) evidence from code-switching or ‘mixed-language’ utterances. Our research questions are:

1. *What aspects of the grammar* are most likely to be affected, or *which linguistic structures* are involved in cross-linguistic interaction in the young bilingual child?
2. What *types of interaction* have been reported? And to what extent does *dominance* (or greater proficiency) in one language account for the effects?
3. What role is played by the *relative accessibility and compatibility of structures*?

3 Morphosyntactic interaction between linguistic systems in early bilingual child production

Our division of the studies into those addressing morphosyntactic effects in single-language utterances and those addressing intraclausal mixing (including intralexical mixing, between the stem and a bound morpheme) will give an idea of the ways in which interaction is manifested.

3.1 Evidence from single-language utterances

Thirty-eight (81%) of the 47 studies in our sample address the issue of morphosyntactic interaction in single-language utterances. Since the number of children in most of these studies is small, we group the studies together by the linguistic structures they analyze, to the extent possible (Tables 1–8, according to topics

Marilyn M. Vihman & Virve-Anneli Vihman

investigated); some studies appear under more than one topic (and some studies also consider mixed-language utterances). The tables indicate the linguistic structures, the language pairs included, the language of the children's community (with the country or region where data were collected, if specified), and the number of children per language pair. Each study is marked for the type of interaction reported, if any: *acceleration*, *delay* – not including delay said to be due to *lack of sufficient experience* with the language in question – or *transfer* (effects other than acceleration or delay). The studies are listed chronologically within each linguistic structure.

Under Dominance, we indicate whether the authors find it to be a factor in the linguistic interaction they identify; N/A indicates that dominance does not figure in the analyses. Relative 'dominance' or 'balance' was established in different ways in the 20 studies that mention it. Eleven studies base their characterization on MLU counts. Kupisch (2007) includes additional measures (type/token ration, upper bound, or longest utterance, number of noun and verb types, number of utterances in 30 minutes). Others also consider aspects of the child's input in each language, based on attendance at school (Austin 2009) or nursery (Blom 2010), parental questionnaire (Fernández Fuertes & Licerias 2010, Licerias et al. 2012) or the child's daily activities with different interlocutors (Qi & Di Biase 2020). Five more studies calculate MLU but do not base their judgment of balance on it. For example, Döpke (1998, 2000) and Hulk & Müller (2000) make use of MLU to establish developmental phases rather than dominance. Mishina-Mori (2020) calculates MLU but uses number of word types to establish dominance. Finally, both Adnyani et al. (2018) and Silva-Corvalán & Montanari (2008) are diary studies that rely on informal observation of the child's experience of the two languages to assess relative dominance. In short, language use by both the children and their interlocutors is variously considered in these studies, with varying degrees of rigor.

Despite differences in methods, analyses and interpretations, 25 of the 38 studies listed in Tables 1–8 (66%) find evidence of interaction in at least one of the structures they analyse.² The studies that address *copula selection or omission* (Table 1) and *object omission* (Table 2) all provide evidence of interaction.

In contrast, of the six studies looking at *inflection or subject-verb agreement* (Table 3), only two report interaction.

Of the remaining studies focusing on inflection, Sinka & Schelleter (1998), in a case study of Latvian-English acquisition, find no cross-linguistic effects,

²Note that although we list studies repeatedly in the tables where more than one language pair was studied, we count the studies only once each in regards to their findings on interaction and the role of dominance.

6 Linguistic interaction in early bilingual development

Table 1: Studies investigating cross-linguistic interaction in **copula selection or omission** in single-language utterances. The community language(s) are marked in bold. We indicate whether the authors see Dominance as a factor affecting the interaction (+ / -); NA indicates that the study does not mention dominance.

Study	Languages (location)	N	Dominance	Type of Interaction
Silva-Corvalán & Montanari (2008)	English Spanish (United States)	1	+	Delay in acquisition of Spanish <i>ser/estar</i>
Fernández Fuertes & Licerás (2010), Licerás et al. (2012)	English Spanish (Spain)	2	-	Acceleration: Less copula omission in English (due to more complex Spanish system)

but provide neither reference to baseline nor monolingual data. Emphasizing the need for evidence of systematic differences to support a claim of interaction, Paradis & Genesee (1996) review studies of monolingual acquisition in French and English that demonstrate relatively earlier mastery of the French system of inflectional marking, due in part to the more central role of inflection in French. Their subsequent analysis of the acquisition of finiteness and agreement in three French-English bilingual children aged 2-3 years shows that they acquire the two systems in much the same way as monolinguals, with the children expressing finite verbs in French before they begin to do so in English (although the bilinguals do produce more finite utterances in English overall than is typical of monolinguals). Hacohen & Schaeffer (2007) compare the Hebrew production of a single Hebrew-English-learning child with five Hebrew monolingual learners. They find an equally low percentage of subject-verb agreement errors in all the children. And Soriente (2014) finds a delay in the acquisition of Italian inflection but considers it the result of insufficient input in that language.

As regards *wh-movement* (Table 4), the two studies that find no interaction – Park-Johnson (2017) and Qi & Di Biase (2020) – both concern early sequential bilingualism.

The children observed in Park-Johnson’s study initially experienced mainly Korean input, not only in the home but also in church and weekend Korean language school. The children began to learn and use English, the language of the

Marilyn M. Vihman & Virve-Anneli Vihman

Table 2: Studies investigating cross-linguistic interaction in **object omission** in single-language utterances.

Study	Languages (location)	N	Dominance	Type of Interaction
Hulk & Müller (2000), Müller & Hulk (2001)	Dutch French (The Netherlands)	1	-	Delay in object provision in Romance languages
	French German (Germany)	1		
	German Italian (Germany)	2		
Yip & Matthews (2000)	Cantonese English (Hong Kong)	1	+	Delay in object provision in English
Mishina-Mori (2020)	English Japanese (United States)	2	-	Delay in object provision in English

larger (American) society, at nursery school. In the case of Qi & Di Biase, the child’s home (in Australia) included monolingual Mandarin-speaking grandparents in addition to the parents and an aunt. The child did not actively begin *producing* English until after six months of silent attendance at a childcare center, beginning at 2;8. In these studies, then, the children were mainly exposed to the home language up to about age 3, when they began to learn and use English in a nursery setting. Such early successive bilingualism may be qualitatively different from bilingual exposure and use from the start (Vihman 1982).

Eight studies addressed *subject provision* (Table 5) in languages that do not require it (and in contexts in which overt pronominal subject provision is pragmatically marked): Six of these found evidence of overprovision of subjects in the children’s pro-drop language, including one study of Spanish-English acquisition (Paradis & Navarro 2003, who deemed their findings inconclusive), while two other studies of Spanish-English acquisition (Licerias et al. 2012, Villa-García

6 Linguistic interaction in early bilingual development

Table 3: Studies investigating cross-linguistic interaction in **inflection**, or **subject-verb agreement** in single-language utterances.

Study	Languages (location)	N	Dominance	Type of Interaction
Paradis & Genesee (1996)	English French (Montreal, Canada)	3	NA	None
Sinka & Schelletter (1998)	Latvian English (England)	1	NA	None
Austin (2007)	Basque Spanish (Spanish Basque country)	20	NA	Delay in Basque ergative use
Hacohen & Schaeffer (2007)	English Hebrew (Israel)	1	-	None
Soriente (2014)	Indonesian Italian (Indonesia)	1	+	None
Adnyani et al. (2018)	German Indonesian (Indonesia)	1	+	Transfer: Vocative + infinitive predicates <i>pro</i> subject + finite verb in German

& Suárez-Palma 2016) did not.³ Blom 2010, finds underprovision of obligatory subjects in Dutch, an effect of Turkish, but only in the case of children for whom Dutch is the weaker language.

Five of the eight papers investigating aspects of *word order* report evidence of cross-linguistic effects, while three do not (Table 6).

Paradis & Genesee (1996) compare the bilingual children’s production in the two languages as regards placement of the negator, which follows the finite verb in French but precedes it in English. Although the data from their three participants is sparse, these investigators do find some evidence of English influence on the children’s French, such as pre-verbal placement of the negator or the use

³Three of the four bilinguals included in Villa-García & Suárez-Palma (2016) were also included in either Paradis & Navarro 2003, or Licerias et al. 2012.

Marilyn M. Vihman & Virve-Anneli Vihman

Table 4: Studies investigating cross-linguistic interaction in **WH-movement** in single-language utterances.

Study	Languages (location)	N	Dominance	Type of Interaction
Yip & Matthews (2000)	Cantonese English (Hong Kong)	1	+	Delay in English WH-fronting
Mishina-Mori (2005)	English Japanese (United States)	2	NA	Delay: Overuse of WH-fronting in Japanese
Soriente (2007)	Italian Indonesian (Indonesia)	1	+	Delay in acquisition of WH-order in Italian
Park-Johnson (2017)	English Korean (United States)	3	NA	None
Qi & Di Biase (2020)	English Mandarin (Australia)	1	-	None
Yip & Matthews (2000)	Cantonese English (Hong Kong)	1	+	Delay in English WH-fronting

of *non* in lieu of *pas* (an apparent calque on English *not*). However, they dismiss ‘these aberrant examples’ as ‘most likely performance errors’ (p. 15).

Sinka & Schelletter (1998) find no evidence of cross-linguistic effects on noun-verb placement in a child learning English and German, in the UK, at 2;0-2;6, while Döpke (1998, 2000) reports a delay in the acquisition of German VP order under the influence of English in four children learning English and German in Australia. Analyzing data from earlier studies of one bilingual Italian-German and one monolingual German child, Hauser-Grüdl et al. (2010) find acceleration in German under Italian influence: The monolingual German learner they observe goes through a phase of using, in main clauses, the verb-final word order pattern found in German subordinate clauses; this persisted to about age 2;8. Thereafter, the child gradually uses more verb-second order (V2) in main clauses.

6 Linguistic interaction in early bilingual development

Table 5: Studies investigating cross-linguistic interaction in **Subject provision** in single-language utterances.

Study	Languages (location)	N	Dominance	Type of Interaction
Paradis & Navarro (2003)	English Spanish (England)	1	NA	Delay: Overprovision of subjects in Spanish
Serratrice et al. (2004)	English Italian (Scotland)	1	NA	Delay: Overprovision of pronominal subjects in Italian
Hacohen & Schaeffer (2007)	English Hebrew (Israel)	1	-	Delay: Overprovision of subjects in Hebrew
Blom (2010)	Dutch Turkish (The Netherlands)	4	+	Delay: Under-provision of subjects in Dutch (when Turkish-dominant)
Haznedar (2010)	English Turkish (Turkey)	1	NA	Delay: Overprovision of subjects in Turkish
Liceras et al. (2012)	English Spanish (Spain)	2	-	None
Schmitz et al. (2012)	German Italian (Germany)	3	NA	Delay: Overprovision of pronominal subjects in German
Villa-García & Suárez-Palma (2016)	French German (Germany)	4	NA	None
	French Italian (France)	1		
	English Spanish (USA, England)	2		
	English-Spanish (Spain)	2		

replace ‘England’ with ‘UK’?

Marilyn M. Vihman & Virve-Anneli Vihman

Table 6: Studies investigating interaction in **word order** in single-language utterances.

Study	Languages (location)	N	Dominance	Type of Interaction
Paradis & Genesee (1996)	English French (Montreal, Canada)	3	NA	None
Döpke (1998)	English German (Australia)	3	-	Delay in acquisition of word order in German VP
Sinka & Schelletter (1998)	English German (England)	1	NA	None
Döpke (2000)	English German (Australia)	4	-	Delay in acquisition of word order in German VP
Hauser-Grüdl et al. (2010)	German Italian (Germany)	1	NA	Acceleration: Earlier adherence to German main-clause V2 order
Adnyani et al. (2018)	German Indonesian (Indonesia)	1	+	Transfer: Use of German VP order in Indonesian
Anderssen & Bentzen (2018)	English Norwegian (Norway)	3	-	Transfer: Overuse of V2 in English
Sağın-Şimşek & Antonova-Ünlü (2021)	English Turkish (England)	4	NA	None
	German Turkish (Germany)	4		
	Russian Turkish (Russia)	4		

6 *Linguistic interaction in early bilingual development*

In contrast, the bilingual child fails to show early use of V-final order, producing V2 from early on instead. ‘Since in Italian, Vfinal word order patterns hardly exist, we take this lack of Vfinal in Lukas to be due to cross-linguistic influence of his Italian L1 onto his German L1.’ (p. 2643)

The four children in each of the three bilingual groups in [Sağın-Şimşek & Antonova-Ünlü \(2021\)](#) showed, in their Turkish production, a developmental trajectory distinct from that of the four Turkish monolinguals. Specifically, all the children make use of the canonical Turkish SOV order in the earliest recordings, but the monolinguals make increasing use of pragmatically based non-canonical orders as they grow older (up to 30% by age 5;5), whereas the heritage-language Turkish learners continue to make almost exclusive use of SOV. The authors argue that a lack of sufficient exposure to Turkish input can best account for the difference; they find direct cross-linguistic influence to be implausible, given that the Russian bilingual learners, who are exposed to more variable word orders, show no better ability to adopt the kind of pragmatic variation Turkish allows than do the children acquiring English or German, with their more fixed verb/object orders.

Root infinitives (RIs, Table 7) involve the delayed acquisition of finite verb marking.⁴ They are the main exception to the general finding of interaction: Here, none of the studies found any such evidence. [Hulk & Müller \(2000\)](#) argue that although Root infinitives satisfy one of the two conditions they posit for cross-linguistic influence (involving the interface between syntax and pragmatics), they fail to fulfill the second condition, with cross-linguistic ‘overlap’ such that ‘the input of one of the adult languages reinforces a misanalysis of RIs (in declarative root clauses) as correct in the other language’ (p. 240). In a close study of one Dutch-French and one German-Italian bilingual child Hulk and Müller find that the rate of use of RIs falls within the range of monolingual children for each of the languages, although they are rare in Italian.

[Unsworth \(2003\)](#) extends [Hulk & Müller](#)’s analysis to an English-German bilingual child, where a cross-linguistic effect might be expected, given the surface similarity between the two languages: Bare verb forms occur in German input as well as in English. Unsworth’s bilingual participant, growing up in Germany, produces too few English bare forms for analysis, but no effect of English on her German is observed with respect to root infinitives: ‘Annie’s German RIs exhibit the same referential properties and contain the same predicate types as her monolingual peers’ (p. 154). Unsworth suggests that the ‘overlap’ condition must mean

⁴We distinguish ‘inflection’ from ‘root infinitives’ based on the terms used in the different studies.

Marilyn M. Vihman & Virve-Anneli Vihman

Table 7: Studies investigating interaction in **Root Infinitives** in single-language utterances.

Study	Languages (location)	N	Dominance	Type of Interaction
Hulk & Müller (2000)	Dutch French (Netherlands)	1	–	None
		1		
	German Italian (Germany)	1		
Unsworth (2003)	English German (UK)	1	NA	None
Austin (2009)	Basque Spanish (Spanish Basque country)	20	–	None

‘overlap between forms in the input of the two languages’, and specifically, partial overlap.⁵ A bilingual child learning German, with its more frequent overtly inflected verbs, could be expected to more easily move beyond the RI stage (i.e., acquire the inflectional markers in English), though Unsworth’s data are insufficient to test the idea. (Compare this with the discussion of the acquisition of finiteness in French and English in Paradis & Genesee 1996.)

In a comparatively large cross-sectional study, Austin (2009) does find more omission of obligatory verbal inflection (carried by auxiliaries) in the Basque of the younger bilinguals (under age 2;8) than in monolinguals, correlated to some extent with lower MLU, but Austin argues against interpreting this as reflecting influence from Spanish. Both the monolingual and the bilingual children had already advanced beyond the ‘root infinitive stage’ in Spanish, where verbal inflection tends to be learned considerably earlier than the equivalent morphology in Basque. Austin draws instead on Gathercole’s (2007) suggestion that a relatively lower amount of input in each of the languages of bilingual children may lead to a temporary delay in some aspects of morphosyntactic acquisition, in comparison with typical monolingual development; a ‘critical mass’ may be needed for the acquisition of some grammatical features. The fact that most of the children showing use of more adult-like morphology in Basque were attending schools

⁵As a reviewer pointed out, in the case of fully overlapping adult structures cross-linguistic effects could not be detected in a child’s production.

6 Linguistic interaction in early bilingual development

where all instruction is provided in Basque (with Spanish taught as a subject) supports this hypothesis; children attending schools where the day is split between instruction in the two languages tended to make use of root infinitives in Basque for a longer period. Parental usage was not monitored. Thus ‘delay’ here, *vis à vis* monolingual expectations, is taken to be the result of the process of bilingual acquisition *per se* (with input split between two languages and therefore reduced in each), rather than being due to any structural characteristic of the competing language.

The studies addressing aspects of *determiner phrases* (DPs, Table 8) are the most heterogeneous in terms of the specific object of analysis.

Kupisch (2007), looking at determiner provision in four children acquiring German and Italian in Germany, found evidence of acceleration in German under the influence of the more accessible determiners in Italian. In contrast, Kupisch et al. (2002), who considered data from both Italian and French children bilingual with German, found later acquisition of gender marking on determiners in the one French child, but ascribed the delay to a lack of sufficient exposure to French.

Hervé & Serratrice (2018) found evidence of acceleration in the acquisition of English determiners under the influence of French, even when French was the weaker language. Cross-linguistic transfer was also observed, from English to French, resulting in article omission, but only in periods of English dominance and with considerable lexical specificity.

Anderssen & Bentzen (2013) and Rodina & Westergaard (2013) analyzed DPs in one and two girls (respectively) learning Norwegian in Norway with English as the home language; they focused on definiteness and gender/declension class, respectively, and found good evidence of interaction in one child, Emma, while in most respects the second child in Rodina and Westergaard’s study resembled the two monolingual children in the study more than she did her younger sister Emma. (Note that in Norwegian definite articles are expressed by inflection rather than by separate determiners, as in the other languages involved here.)

On the other hand, Eichler et al. (2013) failed to find evidence of interaction in gender marking in children learning German (with its three genders) and a Romance language (with two) but did observe some delay in the children’s weaker languages.

The bilingual data in Larrañaga & Guijarro-Fuentes (2013) derive from two children learning Spanish alongside Basque, a language that lacks both gender features and nominal agreement. The bilinguals are found to master gender assignment more slowly than the monolingual Spanish-learning child who serves as a point of comparison. The authors tentatively ascribe the delay to Basque influence on Spanish, but they consider their data too limited for a solid conclusion.

Marilyn M. Vihman & Virve-Anneli Vihman

Table 8: Studies investigating cross-linguistic interaction in single-language utterances: **Determiner phrases**.

Study	Languages (location)	N	Dominance	Type of Interaction
Kupisch et al. (2002)	German Italian (Ger.)	2	-	None
	French German (Ger)	1		
Kupisch (2007)	German Italian (Germany)	4	+	Acceleration: In determiners in German
Chang-Smith (2010)	English Mandarin (Australia)	1	NA	None
Eichler et al. (2013)	French Italian (not specified)	5	+	Delay: Errors in neuter marking in German three-way gender system (due to parallel acquisition of Romance two-way system)
	French German (Germany)	5		
	German Italian (Germany)	7		
	German Spanish (Ger)	2		
	Basque Spanish (Spain)	2		
Larrañaga & Guijarro-Fuentes (2013)	English Norwegian (Norway)	1	-	Delay in acquisition of Spanish gender
Anderssen & Bentzen (2013)	English Norwegian (Norway)	2	NA	Delay in acquisition of Norwegian determiner affix
Rodina & Westergaard (2013)	English French (England)	2	+	Acceleration in Eng. determiner provision Delay in French plural/ generic determ. provision

6 Linguistic interaction in early bilingual development

Finally, Chang-Smith (2010) found no evidence of interaction in the production of classifiers in the single child she observed, acquiring Mandarin alongside English in the United States.

Overall, we find that cross-linguistic interaction is widely reported in the literature on early bilingualism. However, it is not found in all cases, and some authors note that the findings based on small samples are inconclusive. The results are complex, as studies approach the issue from different angles and the evidence may be difficult to evaluate in the absence of baseline or comparative data.

3.2 Evidence from code-switching or “mixed-language utterances”

We initially identified 21 papers investigating *mixed utterances* in bilingual children. However, many of these were not directed at the question of linguistic interaction and failed to specify the morphosyntactic structures involved in any mixing. Those papers were concerned with other issues, such as evidence of children’s ability to separate their languages or to follow adult-like constraints, the children’s language choices, or the extent to which the input may be the source of children’s mixing.

Here we restrict ourselves to the ten studies that include the question discussed above: Which language structures are involved in cross-linguistic interaction in the young bilingual child? Although some of these studies specifically pick out aspects of determiner phrases (DPs) for analysis, most fail to focus on any one structure but instead provide analyses and examples of children’s utterances that do or do not follow the constraints found to apply to much of adult code-switching, most often referencing Myers-Scotton’s (1993) Matrix Language (ML) framework. However, each of the studies can be broadly categorized as primarily addressing one of three aspects of morphosyntax (see Tables 9–11). Interaction between the children’s two languages is necessarily implicated in these studies, with their attention to mixed-language utterances.⁶

The studies are listed chronologically within each table. The **community language(s)** are marked in bold (with study location). We indicate whether the authors see Dominance as a factor affecting the interaction in question (+ / -). An important factor that differs across the studies is the question of compatibility between the two languages with respect to the structures in question; this is indicated in the rightmost column, under Compatible structures, showing whether the structures overlap in the languages in question (+ / -).

⁶Under ‘mixed language utterances’ we include only intraclausal or intralexical switches, in which the grammars of the two languages confront each other (Myers-Scotton 1993).

Marilyn M. Vihman & Virve-Anneli Vihman

As regards the characterization of dominance, five studies (Cantone & Müller 2008, Eichler et al. 2012, Gaskins et al. 2021, Paradis et al. 2000, Yip & Matthews 2016) formally compare the children's MLU in the two languages (although Yip & Matthews 2000 caution that establishing MLU for Cantonese can be problematic); Paradis et al. also consider vocabulary and parental report, while Cantone & Müller additionally include Upper Bound, lexical diversity (numbers of noun and verb types) and number of utterances in half-hour recordings. Soriente (2014) and Vihman (2016, 2018) draw on informal observation of the child's usage and extent of exposure to the two languages. Gawlitzek-Maiwald & Tracy (1996) report dominance based on the children's language preferences (proportion of utterances produced in each language). Bernardini & Schlyter (2004) recruited unbalanced bilingual participants to test their ideas; they base their assessment of dominance on extent of input in each language.

3.2.1 Compatibility of structure

Studies investigating cross-language interaction in *single-language* utterances must necessarily select a linguistic phenomenon which differs across the two languages to identify the interaction. Unlike the studies discussed in Section 3.1, all the studies included in Section 3.2 do by definition provide evidence of some interaction between languages, given the language mixing in the utterances. An additional question that arises here is the effect, on mixing, of the compatibility or equivalence of the relevant structures in the two languages. Note that it has long been claimed that code-switching requires some degree of compatibility (Poplack 1980, Myers-Scotton 1993). As noted above, an early suggestion for predicting where interaction could occur in bilingual development was (partial) overlap between the structures of the two languages (Hulk & Müller 2000); for mixing in adult studies, the point was encapsulated in the 'structural equivalence' constraint (Poplack 1980). Accordingly, we mark the language pairs for compatibility in Tables 9–11. Here we see that about half of the studies investigate intraclausal mixing in structures which are not compatible across the two languages.

We define compatibility as overlap, or similarity, both in linear order and in the type of linguistic elements required for the construction in question. Compatibility is more nuanced than the binary coding suggests, however: For instance, an inflectional paradigm can include areas of overlap (e.g. suffixes are required for third person, present-tense verbs in both English and German) and of non-overlap (first and second person suffixes are required in German but not English;

6 Linguistic interaction in early bilingual development

two languages may both require prenominal articles, but categorize them differently for gender, as with French and German). In this section, we have nevertheless marked the structure under investigation as either compatible or not, as this informs both the research question and the potential for structural interaction. Compatibility again implies partial rather than complete overlap (see fn. 5).

3.2.2 Determiner Phrases

Table 9 gives an overview of the studies in our sample investigating cross-linguistic interaction in determiner phrases in mixed-language utterances. German and Italian share the basic structure Determiner + Noun, with the gender of the noun indexed on the determiner in both languages. Cantone & Müller (2008) find that in 5% of the mixed-language DPs the children encode on the determiner not the gender of the switched noun they produce, but that of the *equivalent noun* in the other language. In other words, the children sometimes access the intended word form in one language but index, on the obligatory determiner, the gender of the equivalent noun in the other language.

We give one example with a German noun in (1), from a balanced bilingual child aged 3;8⁷:

- (1) al schlange
 to.DEF.M snake.F
 (cf., Italian:
 al serpente)
 to.DEF.M snake.M
 ‘to the snake’

Here, then, the DP structures are compatible, but the lexical categorization of gender is not. These errors are more often seen in the productions of children with less balanced bilingualism, but they occur in the stronger as well as the weaker language. While 5% is a small minority of mixed-language DPs, it is unlikely to come from random errors in gender assignment, as the data rarely show errors with nouns which have the same gender in the two languages.

Eichler et al. (2012) also report this minority pattern in their study of 15 children in four different language pairs, with the gender matching the equivalent noun in the child’s other language some 20-25% of the time, considerably more

⁷They do not state how they drew the line between the ‘balanced’ and ‘unbalanced’ bilingual children whose data they present.

Marilyn M. Vihman & Virve-Anneli Vihman

Table 9: Studies investigating cross-linguistic interaction in **determiner phrases** in mixed language utterances. The community language(s) are marked in bold. We indicate whether the authors see Dominance as a factor affecting the interaction (+ / -) and whether the structures under investigation are seen as partially compatible (+ / -).

Study	Languages (location)	N	Dominance	Compatible structures
Cantone & Müller (2008) Eichler et al. (2012)	German Italian (Germany)	4	-	+
	French Italian (not specified)	1	-	+
	French German (Germany)	5		
	German Italian (Germany)	7		
	German Spanish (Germany)	2		
	French Russian (France)	1	-	-
Gaskins et al. (2021)	English German (Germany)	1		+
	English Polish (England)	2		-

often than in Cantone & Müller (2008). Eichler et al. (2012) find that the determiner is always in the language of the verb, which is taken to establish the ML for the clause⁸; only the noun is switched, and the gender is generally that of the switched noun, not that of the verb and determiner, despite its being marked on the determiner. Balanced and unbalanced bilingual children do not differ in this regard, nor does it matter whether the noun comes from the child’s stronger or weaker language (see Eichler et al. 2012, Figure 6: 250).

Finally, Gaskins et al. (2021) include three language pairs, only one of which involves compatible DP structures: English and German both require articles before

⁸Identification of the ML is a major problem for the analysis, especially in two-word combinations. Eichler et al. (2012) find that internal evidence – i.e., the language of the verb – is a more reliable anchor for analysis than the language of the discourse context, typically set by the interlocutor but not necessarily respected by the child. (See detailed discussion in Paradis et al. 2000: 251f.; Bernardini & Schlyter 2004, also address the issue.)

6 *Linguistic interaction in early bilingual development*

the noun in most DPs, whereas Polish and Russian lack articles. These authors find that, although the two children learning Polish and English do not switch elements within the DP, equivalence between structures cannot constitute an absolute constraint, given that the child learning Russian and French does insert both ‘proto-determiners’ and indefinite articles from French before Russian nouns. The study does not report the language of the clause or utterance in which they occur. As the child advances in her knowledge of Russian, the switches disappear. [Gaskins et al. \(2021\)](#) suggest that mastering a language with a greater variety of article types (due to case and gender marking) supports earlier ‘segmentation’, or the distinction between determiner and noun within the DP; this distinction facilitates switching.

3.2.3 **Inflection**

In [Table 10](#) we see the extent of cross-linguistic interaction in inflection in mixed-language utterances in the studies in our sample.

Table 10: Studies investigating cross-linguistic interaction in **inflection** in mixed language utterances.

Study	Languages (location)	N	Dominance	Compatible structures
Paradis et al. (2000)	English French (Montreal, Canada)	15	-	+
Bernardini & Schlyter (2004)	Italian Swedish (Sweden)	1	+	-
	French Swedish (Sweden)	2		
Soriente (2014)	Indonesian Italian (Indonesia)	1	+	-
Vihman (2016)	English Estonian (Estonia)	2	-	+
Vihman (2018)	English Estonian (Estonia)	2	-	-

In a study of 15 French-English bilingual children living in the francophone community of Montreal, [Paradis et al. \(2000\)](#); see also [Paradis & Genesee 1996](#)) set

Marilyn M. Vihman & Virve-Anneli Vihman

out to determine the extent to which the children's code-switching adheres to the structural constraints widely claimed to govern adult switches (Myers-Scotton 1993). Analyses of the children's mixed-language utterances in recordings made at ages 2;0, 2;6, 3;0 and 3;6 also made it possible to test Meisel's (1994) proposal that in the earliest period of word combination linguistic principles would not yet apply, but that once children have developed productive use of functional categories, both single- and mixed-language utterances could be expected to conform to linguistic principles. A developmental difference in the acquisition of verbal markers of tense and agreement between French (early) and English (later) provides a useful testing ground for these ideas.

Use of French system morphemes (e.g., auxiliary verbs, copulas, modals, tense and agreement inflections) in English clauses proved to be relatively frequent (at 20-30%) in the second and third recordings; the more balanced bilinguals produced proportionately more such switches (once their utterances had become more advanced or ambitious), in accord with the 'bootstrapping' hypothesis of Gawlitzek-Maiwald & Tracy (1996). The authors find no evidence of a sharp developmental shift, such as Meisel predicted; Myers-Scotton's system-morpheme constraint was more often observed (82%) than violated across all four sessions for the 15 children. (Note, however, that Paradis et al. determined ML based on the developmental psycholinguistic criterion, excluding mixed utterances from the calculation, rather than on a clause-by-clause basis, as has been done in more recent work.) Paradis et al. conclude that 'it is the children's lexicons that must develop in order to give them the tools to adhere more strictly to the constraint' (2000: 259).

The 'ivy hypothesis' (Bernardini & Schlyter 2004) further elaborates this line of thinking. Focusing on three children with more advanced development in Swedish than in their other language, French or Italian, Bernardini & Schlyter (2004) find that most switching involves clausal structure in Swedish with 'lower' elements taken from the less dominant, less developed language. The authors conclude that 'practically all mixed utterances produced by these children during the period studied were such that the "missing" elements – which represent portions of higher syntactic structure – were "replaced by" their counterparts in the Stronger Language' (p. 65f.). This is the case despite the fact that Swedish, as a V2, non-pro-drop language, has a clausal structure that is incompatible with those of the Romance languages.

The case study Soriente (2014) presents involves a still greater imbalance between the child's two languages, Jakarta Indonesian, the language of the community where he is being brought up, and Italian, his mother's language. The two languages are structurally very different: Italian requires suffixal markers

6 *Linguistic interaction in early bilingual development*

of tense and aspect whereas these concepts are marked lexically, and only optionally, in Indonesian, where the pragmatics of the situation often replace overt marking. Indonesian was the dominant language, used by a full-time caregiver as well as the father. The child was found to first incorporate Indonesian aspectual function words into his Italian, and then to produce Italian word forms that were the rough equivalent of those words, before he began to produce actual inflections in Italian. For example, the earliest-learned Indonesian verbal aspect marker, perfective (s)*udah*, was frequently inserted into Italian clauses, as shown in (2a); later, Italian *basta* ‘enough’ was used to express past tense, taking the place of *udah*, as in (2b).

- (2) a. **udah** volato (2;9)
 PERF flown
 ‘[It] flew away’
 b. Gulli **basta** fare il bagno con Papà.
 G. enough do.INF the bath with Daddy
 ‘Gulli **enough** had a bath with Dad’

Given the relative transparency of the Indonesian system, in addition to the child’s greater exposure to input in that language, the uses the child makes of the ‘scaffolding’ that Indonesian can supply are understandable and in agreement with the bootstrapping idea. The advance from switching to loan translation or calques before the child finally begins to acquire Italian inflections is an interesting further step, one that has seldom been reported.

Vihman (2016, 2018) looks at diary data in two siblings exposed to both Estonian and English in the home and Estonian in the larger community, including preschool. Despite the probable imbalance in input in favor of Estonian, the children mostly spoke in English to one another; the two languages were largely in balance for both. The data, which derive primarily from the younger child at around age three, provide evidence of intense co-activation between the two languages and interaction of the grammatical systems, with use of both bare noun and verb stems and switches in inflectional marking in both languages.

Vihman (2016) focuses on verbal morphology. Although the languages are to some extent compatible here, as both mark tense with suffixes, Estonian has complex morphology, often including stem allomorphy, and generally expresses tense and person with separate suffixal morphemes. In general, these diary data draw on the daily life of the child, while also being avowedly biased toward more complex or puzzling instances, or (as Serratrice 2005, suggests with reference to another diary study) more creative uses. Vihman points up the many types of

6 Linguistic interaction in early bilingual development

Table 11: Studies investigating cross-linguistic interaction in **word order** in mixed language utterances.

Study	Languages (location)	N	Dominance	Compatible structures
Gawlitzek- Maiwald & Tracy (1996)	English German (Germany)	1	+	-
Yip & Matthews (2016)	Cantonese English (Hong Kong)	9	+	+
Vihman (2018)	English Estonian (Estonia)	2	-	-

parallel syntactic formats’ (p. 906), as in *Ich cover michself up* ‘I cover myself up’ (produced at 2;6). Yet mixing also occurs in clauses that require different word orders in the two languages, as in *Cleanst-du dein teeth?* ‘Did you clean your teeth?’ (2;9), where the subject-verb inversion of a main verb observes German but not English clausal structure. The suggestion is that earlier development of constructions in one language allows the child to ‘pool her resources’, ‘taking and combining what is available to her in both languages, in a lexical as well as in a structural sense’ (p. 920). Among her examples of convergence in word order in single-language utterances Vihman (2018) includes production, in English-only clauses, of the inverted object-verb-subject order that occurs in Estonian object-initial syntax. This similarly suggests that grammatical structure may draw on resources from either language. It casts the bilingual utterances in a new light: We have coded ‘compatibility’ with reference to adult structures, but that need not constrain code-switching in early bilingual productions.

Yip & Matthews (2016) address ‘code mixing’ in children learning Cantonese and English in Hong Kong, where switching between the two languages is the community practice. They point out that the congruence in word order between the two languages facilitates switching; this includes the analytic nature of both grammars. For all nine children switching occurs most frequently in the ‘Cantonese context’ (i.e. when the person recording them is using Cantonese), although most of the children also mix in the English context and most show considerable variation across the one- to two-year span of the recordings. In English utterances the addition of a Cantonese sentence-final particle (glossed SFP) is the most common switch, as in (4):

Marilyn M. Vihman & Virve-Anneli Vihman

- (4) You tidy up laa1⁹
 you tidy up SFP
 ‘You tidy up.’ (3;6;8, English context)

The children may feel that something is missing in their English utterances in the absence of such completive particles. The authors note that these particles are also switched by adults in Singapore English, for example.

Unlike what has been found in European languages (e.g., German/ Italian: *Cantone* 2007: 173), verbs are the part of speech most often switched, typically from English into Cantonese. In these cases Cantonese aspect markers (ASP) are added to the English verbs, as shown in (5).

- (5) dim2gaai2 **turn** zo2?
 why turn ASP
 ‘Why did it turn?’ (3;4;18, Cantonese context)

Bare stems in Cantonese are also mixed into English utterances, but the addition of English tense-aspect morphology is rare.

Another frequent type of switch, shown in (6), involves English verb particles, sometimes with Cantonese material intervening between the verb and its particle:

- (6) a. Dim2 gaai2 lei5 **throw** ni1 go3 **away** ge3?
 how come you throw this CL¹⁰ away SFP
 ‘Why are you throwing this away?’ (3;6, Cantonese context)
 b. Ji1 zek3 **slide** m4 dou2 **down**
 this CL slide not succeed down
 ‘(With) these (shoes on) you can’t slide down.’ (2;6, CC)

Yip & Matthews see these split-verb switches, unattested in adult code switching, as reflecting ‘cross-linguistic influence, which takes place between the developing grammars, independent of code-mixing’ (p. 12). In other words, above and beyond the fact of producing a mixed-language utterance, they see the insertion of a Cantonese constituent between the English verb and participle as an effect of English grammatical structure. The authors also find evidence of the English split verb-particle structure in the children’s single-language utterances:

⁹The numbers following Cantonese words mark the lexical tones.

¹⁰CL is presumably ‘classifier’, but Yip and Matthews do not spell it out.

6 Linguistic interaction in early bilingual development

The code-mixed cases...are thus congruent with the children's developmental grammar for Cantonese, if not with the target grammar of Cantonese (p. 12).

This point is echoed in [Vihman \(2018\)](#), in relation to the convergence found in the bilingual siblings' single-language structures.

4 Discussion and conclusions

In their meta-analysis of experimental studies with older children than those included in the studies we review here, [Van Dijk et al. \(2022\)](#) find that interaction between the languages of a bilingual child neither diminishes nor increases with age; they conclude that cross-linguistic influence is 'part and parcel of bilingual development' (p. 902). In our overview of studies of emergent morphosyntax as observed in children aged about 2 to 4 years, we have similarly found that most studies report evidence of linguistic interaction, and that this affects all but one of the structures investigated. In contrast with the findings of Van Dijk et al., however, many of these studies report a decrease in cross-linguistic effects as the children gain mastery of the two languages. Furthermore, a delay in the acquisition of certain structures, compared to monolinguals, often appears to be due to reduced exposure to one of the languages, or of reduced opportunities for its use – due, in other words, to the bilingual experience itself rather than to any specific structural effects. This may be an effect of bilingual language use that is particular to the earliest period of grammatical development.

[Van Dijk et al. \(2022\)](#) also find that neither surface overlap between syntactic structures nor language domain (such as the interface between syntax and pragmatics) are significant predictors of interaction, contrary to much of the early literature. [Serratrice \(2013\)](#), in a qualitative review of the literature to that date, also found counterevidence to the hypothesis that only overlapping structures or interfaces can account for the linguistic interaction identified in the speech of bilingual children. She cites evidence and arguments to suggest that the nature of bilingual processing and use may help to better account for the full range of cross-linguistic influence observed in younger children, both in small-N studies and in larger studies that include tests of comprehension. In other words, it is suggested here again that the bilingual experience itself leads to cross-linguistic effects in children, as it does in adults ([Kroll & Bialystok 2013](#)).

RQ1: *Which linguistic structures are involved in cross-linguistic interaction in the young bilingual child?*

Marilyn M. Vihman & Virve-Anneli Vihman

Our first research question concerned the aspects of grammatical systems likely to be affected by cross-linguistic interaction. To address this, we looked at 47 studies of early bilingual production and attempted to establish which structures are involved in either suspected cross-linguistic influence or inclusion of surface elements of both languages in a single utterance (code-switching). Overall, we found that most of the structures we considered were involved in bilingual interaction of one kind or another. Those structures cover most aspects of the grammatical systems acquired in the first years of language use: determiner use in nominal systems, inclusion of copulas in predicates, subject and object provision, inflection in both noun and verb systems, wh-movement and word order. Accordingly, the basic finding is that almost every aspect of the early developing grammar is open to the influence of bilingual interaction.

RQ2a: What *types of interaction* have been reported?

Tables 1-8 identify findings of acceleration (three structures and four studies), delay (six structures, 20 studies) and transfer (two structures, three studies). These results cannot be deemed definitive, given the small numbers of children involved and the various complications affecting analysis of these few cases (issues around the extent of input and practice the children may have experienced in one of the languages, above all, but also the indisputable effect of individual differences in learning rates, strategy, and so on). However, the overview is certainly indicative of the kinds of interaction that may be found between languages that contrast with respect to a range of morphosyntactic structures. We hope this overview of the literature to date will encourage further systematic study, including studies of larger numbers of younger children learning language pairs that differ in selected ways, to supplement the experimental studies of older children.

RQ2b: *And to what extent does dominance in one language account for interaction effects?*

In [Van Dijk et al. \(2022\)](#), language dominance proved to be the single most significant predictor of interaction – but only when defined in terms of the *relation of the language tested experimentally to the ‘societal language’*, or the language of the community in which the child is living. That is, there was more evidence of interaction in children tested in a language other than that of the community – the ‘weaker’ language, by that definition – although some effects were bidirectional. When dominance was instead defined and measured in ways particular to each study (for these older children the measures included amount of exposure or use, size of lexicon and fluency ratings by caretakers), dominance was not a significant predictor. Van Dijk et al. interpret their failure to find a significant effect in this analysis as likely due to differences in the ways the authors of various studies arrived at their categorization. In the case of children under age

6 *Linguistic interaction in early bilingual development*

four, who may or may not have been regularly exposed to a range of speakers for each of their two languages, the societal language is likely to be less directly connected to relative knowledge or proficiency.

In the literature on younger children, where the issue has long been a subject of debate, dominance is generally measured by Mean Length of Utterance (MLU) in each language, although numerous other measures are used in the studies examined here as well, as discussed in 3.1 and 3.2. Of the 29 studies that mention it, 13 see dominance as a factor. Note that in their account of gender assignment in determiners, [Eichler et al. \(2013\)](#) failed to observe cross-linguistic effects but saw dominance as interacting with the overall ranking of languages in terms of children's accuracy in gender assignment.

On the other hand, at least three studies ([Anderssen & Bentzen 2018](#), [Hulk & Müller 2000](#), [Müller & Hulk 2001](#)) specifically argue that only internal structural issues are at play; they judge that dominance cannot be a key factor, since they find interaction effects to be bidirectional (cf. also [Adnyani et al. 2018](#), who also find bidirectional effects, but with more instances of influence from the dominant to the non-dominant language than vice versa). Given the differences in definition and measurement of dominance across the studies and language pairs we have considered, we can draw no solid conclusions. Ultimately, the effect of dominance may depend more on individual differences, and the shifting strength of a child's proficiency in each language, than on any one bilingual combination or situation.

It is worth noting that the linguistic experiences of children sharing a language pair, even in the same community, may be quite different, depending not only on differences in family language policy and usage but also on the individual child's rate of language development and overall maturation. In general, these studies point up the importance of individual differences, particularly in the early years, when children acquire morphosyntactic knowledge at very uneven rates; comparisons of longitudinal studies of one or two children make this particularly evident ([Serratrice 2005](#)). At older ages, developmental differences begin to even out and the societal (or school) language typically becomes the stronger one.

A nice example of the difficulty of drawing broad conclusions from studies of one or two children is provided by [Anderssen & Bentzen \(2018\)](#), two of whose participants were siblings, 10 years apart in age (but followed over the same early developmental period; cf. also [Rodina & Westergaard 2013](#)): The older child showed little cross-linguistic interaction, whereas the younger child, said to be well balanced in her two languages, showed considerable interaction in both directions. For example, Emma showed the influence of Norwegian in her question formation, as in example (7):

Marilyn M. Vihman & Virve-Anneli Vihman

(7) *Drive daddy me to **barnehage**?*

‘Will/did daddy drive me to nursery?’

(Anderssen & Bentzen 2018: 8)

Emma seemed to find language learning more challenging than her older sister had; she may also have had more ‘tolerance for variability’ (Kamhi, Kamhi et al. 1984), or in other words, more willingness to produce target-like and non-target-like structures in parallel.

RQ3: *What role is played by the relative accessibility and compatibility of structures?*

Few of the studies reviewed here mention the possible role of the relative accessibility of structures in the children’s two languages. Two studies – Kupisch (2007) and Hervé & Serratrice (2018) – report accelerated learning under the influence of the child’s other language, but in an earlier study Kupisch et al. (2002) found no interaction between German and a Romance language with respect to acquisition of the same structure (determiner phrases; cf. also Eichler et al. 2012). Both Kupisch et al. (2002) and Austin (2009) find that a lack of sufficient exposure to one of the languages is the more plausible explanation for the effects they identify.

Recall that among the conditions that Hulk & Müller proposed as being necessary for syntactic cross-linguistic influence to occur was ‘a certain overlap of the two systems at the surface level’ (2000: 228). This requirement has since been somewhat reformulated but not overturned. The overlap refers to (partial) similarity in input structures in the child’s two languages (Unsworth 2003). More specifically, Hulk & Müller proposed that an ambiguous structure that allows for more than one syntactic interpretation in language A may correspond to a structure with only one analysis in language B; cross-linguistic influence occurs when the child takes the analysis in language B as evidence for language A. The studies we review investigate structures with various types of (partial) overlap, including similar elements with different word order (e.g. with wh-movement), similar word order but variation in which linguistic categories are expressed (e.g. gender on determiners), similar inflectional categories but differences in distribution (e.g. verb inflection), and structures with optional expression in one language but not the other (e.g. argument omission).

The structures in which cross-linguistic influence has been documented suggest that it is not limited to those with differences in the analyses available in each language, as Hulk & Müller’s formulation would have it. From Tables 9-11 we see that many structures that lack structural compatibility nevertheless appear in bilingual children’s mixed-language utterances. Soriente’s (2014) exam-

6 *Linguistic interaction in early bilingual development*

ple of the development of aspect marking, lexical in Indonesian and inflectional in Italian, provides a good case in point: The child made use of Indonesian lexical aspect to scaffold the development of aspectual marking in Italian. There is ‘overlap’ only in the fact of overt expression of aspect in both languages. The overrepresentation, in the literature, of European languages (Kidd & Garcia 2022; Yip & Matthews 2022), which tend to be both genetically and typologically similar (see Haspelmath 2001), may have led to a bias in the evidence with regards to the limits of cross-linguistic influence, giving greater weight to mutual compatibility of structures. To more fully test the effects of structural differences in the input we need more studies of both monolingual and bilingual development, in a wider range of languages representing greater typological diversity.

As the abstract categories of the target grammars cannot be said to be fully available to children (e.g., Ambridge 2020), it is unclear how similarity of structure may operate in children’s bilingual production and developing grammars. Experimental and corpus studies of syntactic priming with adults (e.g. Kootstra et al. 2012, Travis et al. 2017) have shown that representations of syntactic structure may be mapped across languages, but that this requires similarity of both constituent order and syntactic structure. Whereas linear order is undeniably available to children, it is unclear to what extent the more abstract syntactic similarity of constructions may influence their usage, or at what age (for syntactic priming in monolingual children see, for example, Rowland et al. 2012, Peter et al. 2015, Branigan & McLean 2016). Some studies have found cross-language effects of structural priming in children (e.g. Hervé et al. 2016, Vasilyeva et al. 2010, Wolleb et al. 2018), but more work is needed to disentangle the interaction of cognitive and linguistic development with cross-linguistic influence.

Children’s developing grammars are a moving target: Children are continuously (i) reshaping their grammatical knowledge, (ii) advancing in their ability to abstract structure out of learned linguistic material, and (iii) developing their ability to produce structures accordingly. Hence, it may be impossible to define what kind of overlap or how much of it there should be between languages to trigger interaction; the overlap itself, as represented in the child’s mind, may undergo continual shifting.

5 **Future directions**

The attempt to define children’s productions, especially utterances with code-switching, according to target language may be misguided. Children’s grammars are in flux, both developmentally and through ongoing interaction with the other

Marilyn M. Vihman & Virve-Anneli Vihman

language. Hence, it is not always appropriate (or even possible) to define a matrix language according to the adult grammar (Vihman 2018) or to assume a target structure for children's productions (Yip & Matthews 2016). Nevertheless, the decades of fruitful research included in our review have resulted in a sizable body of evidence that language interaction does take place, in a wide range of structures and language pairs.

The approach of translanguaging studies (e.g. Wei 2018, Otheguy et al. 2015, Balam 2021) is congenial to the view that children make use of their full linguistic repertoires, often without drawing clear lines between their named languages. The translanguaging literature, drawing on a wider array of bilingual speakers – especially older children in educational contexts – provides a useful reminder that bilinguals' linguistic resources may amount to more than the sum of their parts. That is, use of their languages in combination may be taken to increase expressive capacity. The 'moving target' of emerging grammars mentioned above can be seen to continue to develop as children enter school, while the relative dominance of their languages, as well as the repertoire available to them, may shift. We hope that future research will investigate that trajectory, beginning with the early years, when children have yet to gain a full grasp of either grammatical system, through to a context where they have become proficient and educated in one language and may experience a range of attitudes toward the other, in and out of school and in the home.

Bearing in mind the dynamics of emergent grammatical structures and the shifts in the extent or type of exposure to and opportunities for use of each language over the course of development, we cannot with any confidence pin down the developmental points most likely to reflect interaction between the child's languages, intriguing though that issue might be. Recall the conclusion of Paradis et al. (2000), that it is lexical advance that paces the development of morphosyntactic structure – a conclusion that aligns well with the long-standing views and findings of Elizabeth Bates and her colleagues with respect to monolingual acquisition (cf. Bates & Goodman 1999).

Both generative and Usage-Based accounts have grappled with Root infinitives in monolingual children's language development because the data are hard to reconcile with the theoretical predictions. It is perhaps unsurprising that this is the only linguistic domain in which none of the studies we reviewed report cross-linguistic interaction. As various factors may be at play in the RI conundrum, an approach that looks closely at the structural differences in the languages, as well as the frequency of occurrence in the input of the syntactic structures in question, the word forms themselves and collocational information, would considerably advance our understanding of optional infinitive usage in bilingual development.

6 Linguistic interaction in early bilingual development

Freudenthal et al. (2024) provide a promising learning model that could readily be tested with bilingual children.

Another factor worth investigating in future research is the question of patterns of use in the wider community (see Balam et al. 2021 and van Osch et al. 2023). The extent to which the level and form of bilingualism in the community affects a child's usage is underexplored in general, but particularly in studies of younger children. The effects of community use may differ by child, by age and by caregiving context. Attitudes toward each language and toward language mixing, the broader prestige of the languages, and the extent to which they are used in similar contexts by the same interlocutors may all play a role in children's acquisition of the systems.

A fundamental unresolved issue underlying any aspect of cross-linguistic interaction, whether in single- or mixed-language utterances, is that of children's memory of, or access to, the speech patterns they have heard and practiced, or in other words, the nature of their linguistic representations. We must assume, based on the evidence we have reviewed here, that those representations are variable, not only over developmental time but also over moments of utterance; in addition, we can assume that children experience fluctuations in the relative control available for shaping their utterances to fit these representations, that is, in their control over the process of speech production. Theoretical models developed to account for the intricacies of adult grammar need to incorporate evidence from both language acquisition and bilingual processing. As the increasing pace of studies in this area suggests, the bilingual child can provide an intriguing window into the 'inner life' of language in development.

References

- Adnyani, Ni Luh Putu Sri, Ni Luh Sutjiati Beratha, I Wayan Pastika & I Nyoman Suparwa. 2018. The development of verbal morphology and word order in an Indonesian-German bilingual child: A case study. *Topics in Linguistics* 19(1). 33–53. DOI: [10.2478/topling-2018-0003](https://doi.org/10.2478/topling-2018-0003).
- Ambridge, Ben. 2020. Against stored abstractions: A radical exemplar model of language acquisition. *First Language* 40(5–6). 509–559. DOI: [10.1177/0142723719869731](https://doi.org/10.1177/0142723719869731).
- Anderssen, Merete & Kristine Bentzen. 2013. Cross-linguistic influence outside the syntax-pragmatics interface: A case study of the acquisition of definiteness. *Studia Linguistica* 67(1). 82–100. DOI: [10.1111/stul.12011](https://doi.org/10.1111/stul.12011).

Marilyn M. Vihman & Virve-Anneli Vihman

- Anderssen, Merete & Kristine Bentzen. 2018. Different outcomes in the acquisition of residual V2 and *Do*-support in three Norwegian-English bilinguals: Cross-linguistic influence, dominance and structural ambiguity. *Frontiers in Psychology* 9. DOI: [10.3389/fpsyg.2018.02130](https://doi.org/10.3389/fpsyg.2018.02130).
- Austin, Jennifer. 2007. Grammatical interference and the acquisition of ergative case in bilingual children learning Basque and Spanish. *Bilingualism: Language and Cognition* 10(3). 315–331. DOI: [10.1017/S1366728907003094](https://doi.org/10.1017/S1366728907003094).
- Austin, Jennifer. 2009. Delay, interference and bilingual development: The acquisition of verbal morphology in children learning Basque and Spanish. *International Journal of Bilingualism* 13(4). 447–479. DOI: [10.1177/1367006909353234](https://doi.org/10.1177/1367006909353234).
- Balam, Osmer. 2021. Beyond differences and similarities in codeswitching and translanguaging research. *Belgian Journal of Linguistics* 35(1). 76–103. DOI: [10.1075/bjl.00065.bal](https://doi.org/10.1075/bjl.00065.bal).
- Balam, Osmer, Usha Lakshmanan & María del Carmen Parafita Couto. 2021. Gender assignment strategies among simultaneous Spanish/English bilingual children from Miami, Florida. *Studies in Hispanic and Lusophone Linguistics* 14(2). 241–280. DOI: [10.1515/shll-2021-2045](https://doi.org/10.1515/shll-2021-2045).
- Bates, Elizabeth & Judith C. Goodman. 1999. On the emergence of grammar from the lexicon. In Brian MacWhinney (ed.), *The emergence of language*, 29–79. Mahwah: Lawrence Erlbaum Associates.
- Baum, Shari & Debra Titone. 2014. Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics* 35(5). 857–894. DOI: [10.1017/S0142716414000174](https://doi.org/10.1017/S0142716414000174).
- Bernardini, Petra & Suzanne Schlyter. 2004. Growing syntactic structure and code-mixing in the weaker language: The Ivy Hypothesis. *Bilingualism: Language and Cognition* 7(1). 49–69. DOI: [10.1017/S1366728904001270](https://doi.org/10.1017/S1366728904001270).
- Blom, Elma. 2010. Effects of input on the early grammatical development of bilingual children. *International Journal of Bilingualism* 14(4). 422–446. DOI: [10.1177/1367006910370917](https://doi.org/10.1177/1367006910370917).
- Bosch, Laura, Melània Figueras, Maria Teixidó & Marta Ramon-Casas. 2014. Rapid gains in segmenting fluent speech when words match the rhythmic unit: Evidence from infants acquiring syllable-timed languages. *Frontiers in Psychology* 4. DOI: [10.3389/fpsyg.2013.00106](https://doi.org/10.3389/fpsyg.2013.00106).
- Branigan, Holly P. & Janet F. McLean. 2016. What children learn from adults' utterances: An ephemeral lexical boost and persistent syntactic priming in adult-child dialogue. *Journal of Memory and Language*. New Approaches to Structural Priming 91. 141–157. DOI: [10.1016/j.jml.2016.02.002](https://doi.org/10.1016/j.jml.2016.02.002).

6 Linguistic interaction in early bilingual development

- Byers-Heinlein, Krista. 2013. Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism: Language and Cognition* 16(1). 32–48. DOI: [10.1017/S1366728912000120](https://doi.org/10.1017/S1366728912000120).
- Byers-Heinlein, Krista. 2014. Languages as categories: Reframing the “one language or two” question in early bilingual development. *Language Learning* 64(s2). 184–201. DOI: [10.1111/lang.12055](https://doi.org/10.1111/lang.12055).
- Byers-Heinlein, Krista, Tracey C. Burns & Janet F. Werker. 2010. The Roots of Bilingualism in Newborns. *Psychological Science* 21(3). 343–348. DOI: [10.1177/0956797609360758](https://doi.org/10.1177/0956797609360758).
- Cantone, Katja Francesca. 2007. *Code-switching in bilingual children*, vol. 296. Dordrecht: Springer.
- Cantone, Katja Francesca & Natascha Müller. 2008. Un nase or una nase? What gender marking within switched DPs reveals about the architecture of the bilingual language faculty. *Lingua*. Formal syntactic approaches to bilingual code-switching 118(6). 810–826. DOI: [10.1016/j.lingua.2007.05.007](https://doi.org/10.1016/j.lingua.2007.05.007).
- Chang-Smith, Meiyun. 2010. Developmental pathways for first language acquisition of Mandarin nominal expressions: Comparing monolingual with simultaneous Mandarin-English bilingual children. *International Journal of Bilingualism* 14(1). 11–35.
- Comeau, Liane, Fred Genesee & Lindsay Lapaquette. 2003. The Modeling Hypothesis and child bilingual codemixing. *International Journal of Bilingualism* 7(2). 113–126. DOI: [10.1177/13670069030070020101](https://doi.org/10.1177/13670069030070020101).
- De Houwer, Annick. 1995. Bilingual Language Acquisition. In Paul Fletcher & Brian MacWhinney (eds.), *The handbook of child language*, 219–250. London: Blackwell. DOI: [10.1111/b.9780631203124.1996.00009.x](https://doi.org/10.1111/b.9780631203124.1996.00009.x).
- De Houwer, Annick. 2005. Early bilingual acquisition. In Judith F. Kroll & Annette M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, 30–48. Oxford: Oxford University Press.
- Deuchar, Margaret. 2013. Code Switching. In Carol A. Chapelle (ed.), *The encyclopedia of applied linguistics*. Malden: Blackwell. DOI: [10.1002/9781405198431.wbeal0142](https://doi.org/10.1002/9781405198431.wbeal0142).
- Deuchar, Margaret. 2020. Code-switching in linguistics: A position paper. *Languages* 5(2). 22. DOI: [10.3390/languages5020022](https://doi.org/10.3390/languages5020022).
- Deuchar, Margaret & Suzanne Quay. 2000. *Bilingual acquisition: Theoretical implications of a case study*. Oxford: Oxford University Press.
- Döpke, Susanne. 1992. *One parent, one language*. Amsterdam: John Benjamins.
- Döpke, Susanne. 1998. Competing language structures: The acquisition of verb placement by bilingual German-English children. *Journal of Child Language* 25(3). 555–584. DOI: [10.1017/S0305000998003584](https://doi.org/10.1017/S0305000998003584).

Marilyn M. Vihman & Virve-Anneli Vihman

- Döpke, Susanne. 2000. Generation of and retraction from cross-linguistically motivated structures in bilingual first language acquisition. *Bilingualism: Language and Cognition* 3(3). 209–226. DOI: [10.1017/S1366728900000341](https://doi.org/10.1017/S1366728900000341).
- Eichler, Nadine, Malin Hager & Natascha Müller. 2012. Code-switching within determiner phrases in bilingual children: French, Italian, Spanish and German. *Zeitschrift für französische Sprache und Literatur* 122(3). 227–258.
- Eichler, Nadine, Veronika Jansen & Natascha Müller. 2013. Gender acquisition in bilingual children: French–German, Italian–German, Spanish–German and Italian–French. *International Journal of Bilingualism* 17(5). 550–572. DOI: [10.1177/1367006911435719](https://doi.org/10.1177/1367006911435719).
- Fernández Fuertes, Raquel & Juana M. Liceras. 2010. Copula omission in the English developing grammar of English/Spanish bilingual children. *International Journal of Bilingual Education and Bilingualism* 13(5). 525–551. DOI: [10.1080/13670050.2010.488285](https://doi.org/10.1080/13670050.2010.488285).
- Freudenthal, Daniel, Fernand Gobet & Julian M. Pine. 2024. MOSAIC+: A crosslinguistic model of verb-marking errors in typically developing children and children with developmental language disorder. *Language Learning* 74(1). 111–145. DOI: [10.1111/lang.12580](https://doi.org/10.1111/lang.12580).
- Gaskins, Dorota, Oksana Bailleul, Anne Werner & Antje Endesfelder Quick. 2021. A crosslinguistic study of child code-switching within the noun phrase: A usage-based perspective. *Languages* 6(1). 1–29. DOI: [10.3390/languages6010029](https://doi.org/10.3390/languages6010029).
- Gaskins, Dorota, Antje Endesfelder Quick, Anna Verschik & Ad Backus. 2022. Usage-based approaches to child code-switching: State of the art and ways forward. *Cognitive Development* 64. 1–12. DOI: [10.1016/j.cogdev.2022.101269](https://doi.org/10.1016/j.cogdev.2022.101269).
- Gathercole, Virginia C. M. 2007. Miami and North Wales, so far and yet so near: A constructivist account of morphosyntactic development in bilingual children. *International Journal of Bilingual Education and Bilingualism* 10(3). 224–247. DOI: [10.2167/beb442.0](https://doi.org/10.2167/beb442.0).
- Gawlitzeck-Maiwald, Ira & Rosemary Tracy. 1996. Bilingual bootstrapping. *Linguistics* 34(5). 901–926. DOI: [10.1515/ling.1996.34.5.901](https://doi.org/10.1515/ling.1996.34.5.901).
- Genesee, Fred. 1989. Early bilingual development: One language or two? *Journal of Child Language* 16(1). 161–179. DOI: [10.1017/S0305000900013490](https://doi.org/10.1017/S0305000900013490).
- Genesee, Fred, Elena Nicoladis & Johanne Paradis. 1995. Language differentiation in early bilingual development. *Journal of Child Language* 22(3). 611–631. DOI: [10.1017/S0305000900009971](https://doi.org/10.1017/S0305000900009971).
- Goodz, Naomi Singerman. 1989. Parental language mixing in bilingual families. *Infant Mental Health Journal* 10(1). 25–44. DOI: [10.1002/1097-0355\(198921\)10:1<25::AID-IMHJ2280100104>3.0.CO;2-R](https://doi.org/10.1002/1097-0355(198921)10:1<25::AID-IMHJ2280100104>3.0.CO;2-R).

6 Linguistic interaction in early bilingual development

- Grosjean, François. 1998. Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1(2). 131–149. DOI: [10 . 1017 / S136672899800025X](https://doi.org/10.1017/S136672899800025X).
- Hacohen, Aviya & Jeannette Schaeffer. 2007. Subject realization in early Hebrew/English bilingual acquisition: The role of crosslinguistic influence. *Bilingualism: Language and Cognition* 10(3). 333–344. DOI: [10 . 1017 / S1366728907003100](https://doi.org/10.1017/S1366728907003100).
- Haspelmath, Martin. 2001. The European linguistic area: Standard Average European. In Martin Haspelmath (ed.), *Language typology and language universals*, vol. 2 (Handbücher zur Sprach- und Kommunikationswissenschaft 20), 1492–1510. Berlin: De Gruyter. DOI: [10.1515/9783110194265-044](https://doi.org/10.1515/9783110194265-044).
- Hauser-Grüdl, Nicole, Lastenia Arencibia Guerra, Franziska Witzmann, Estelle Leray & Natascha Müller. 2010. Cross-linguistic influence in bilingual children: Can input frequency account for it? *Lingua* 120(11). Frequency and Language Development, ed. Katrin Schmitz, 2638–2650. DOI: [10.1016/j.lingua.2010.06.008](https://doi.org/10.1016/j.lingua.2010.06.008).
- Haznedar, Belma. 2010. Transfer at the syntax–pragmatics interface: Pronominal subjects in bilingual Turkish. *Second Language Research* 26(3). 355–378.
- Hervé, Coralie & Ludovica Serratrice. 2018. The development of determiners in the context of French–English bilingualism: A study of cross-linguistic influence. *Journal of Child Language* 45(3). 767–787. DOI: [10.1017 / S0305000917000459](https://doi.org/10.1017/S0305000917000459).
- Hervé, Coralie, Ludovica Serratrice & Martin Corley. 2016. Dislocations in French–English bilingual children: An elicitation study. *Bilingualism: Language and Cognition* 19(5). 987–1000. DOI: [10.1017/S1366728915000401](https://doi.org/10.1017/S1366728915000401).
- Hulk, Aafke & Natascha Müller. 2000. Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition* 3(3). 227–244. DOI: [10.1017/S1366728900000353](https://doi.org/10.1017/S1366728900000353).
- Kamhi, Alan G., Hugh W. Catts & Michelle K. Davis. 1984. Management of sentence production demands. *Journal of Speech, Language, and Hearing Research* 27(3). 329–338. DOI: [10.1044/jshr.2703.329](https://doi.org/10.1044/jshr.2703.329).
- Kidd, Evan & Rowena Garcia. 2022. How diverse is child language acquisition research? *First Language* 42(6). 703–735. DOI: [10.1177/01427237211066405](https://doi.org/10.1177/01427237211066405).
- Kootstra, Gerrit Jan, Janet G. Van Hell & Ton Dijkstra. 2012. Priming of code-switches in sentences: The role of lexical repetition, cognates, and language proficiency. *Bilingualism: Language and Cognition* 15(4). 797–819. DOI: [10.1017 / S136672891100068X](https://doi.org/10.1017/S136672891100068X).

Marilyn M. Vihman & Virve-Anneli Vihman

- Kroll, Judith F. & Ellen Bialystok. 2013. Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology* 25(5). 497–514. DOI: [10.1080/20445911.2013.799170](https://doi.org/10.1080/20445911.2013.799170).
- Kroll, Judith F. & Annette M. B. De Groot (eds.). 2005. *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press.
- Kupisch, Tanja. 2007. Determiners in bilingual German–Italian children: What they tell us about the relation between language influence and language dominance. *Bilingualism: Language and Cognition* 10(1). 57–78. DOI: [10.1017/S1366728906002823](https://doi.org/10.1017/S1366728906002823).
- Kupisch, Tanja, Natascha Müller & Katja Francesca Cantone. 2002. Gender in monolingual and bilingual first language acquisition: Comparing Italian and French. *Lingue e linguaggio* 1. 107–150. DOI: [10.1418/7559](https://doi.org/10.1418/7559).
- Lanvers, Ursula. 2001. Language alternation in infant bilinguals: A developmental approach to codeswitching. *International Journal of Bilingualism* 5(4). 437–464. DOI: [10.1177/13670069010050040301](https://doi.org/10.1177/13670069010050040301).
- Lanza, Elizabeth. 1997a. Language contact in bilingual two-year-olds and code-switching: Language encounters of a different kind? *International Journal of Bilingualism* 1(2). 135–162. DOI: [10.1177/136700699700100203](https://doi.org/10.1177/136700699700100203).
- Lanza, Elizabeth. 1997b. *Language mixing in infant bilingualism: A sociolinguistic perspective*. Oxford: Clarendon.
- Larrañaga, Pilar & Pedro Guijarro-Fuentes. 2013. The linguistic competence of early Basque–Spanish bilingual children and a Spanish monolingual child 17(5), 577–601. *International Journal of Bilingualism* 17(5). 577–601. DOI: [10.1177/1367006911435704](https://doi.org/10.1177/1367006911435704).
- Liceras, Juana M., Raquel Fernández Fuertes & Anahí Alba de la Fuente. 2012. Overt subjects and copula omission in the Spanish and the English grammar of English–Spanish bilinguals: On the locus and directionality of interlinguistic influence. *First Language* 32(1–2). 88–115. DOI: [10.1177/0142723711403980](https://doi.org/10.1177/0142723711403980).
- Mayer, Mercer. 1969. *Frog, where are you?* New York: Dial Press.
- Meisel, Jürgen M. 1994. Code-switching in young bilingual children: The acquisition of grammatical constraints. *Studies in Second Language Acquisition* 16(4). 413–439. DOI: [10.1017/S0272263100013449](https://doi.org/10.1017/S0272263100013449).
- Meng, Hairong & Tadao Miyamoto. 2012. Input and output in code switching: A case study of a Japanese–Chinese bilingual infant. *International Journal of Bilingual Education and Bilingualism* 15(4). 393–415. DOI: [10.1080/13670050.2012.665826](https://doi.org/10.1080/13670050.2012.665826).
- Mishina-Mori, Satomi. 2005. Autonomous and interdependent development of two language systems in Japanese/English simultaneous bilinguals: Evi-

6 Linguistic interaction in early bilingual development

- dence from question formation. *First Language* 25(3). 291–315. DOI: [10.1177/0142723705052560](https://doi.org/10.1177/0142723705052560).
- Mishina-Mori, Satomi. 2020. Cross-linguistic influence in the use of objects in Japanese/English simultaneous bilingual acquisition. *International Journal of Bilingualism* 24(2). 319–338.
- Müller, Natascha & Aafke Hulk. 2001. Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition* 4(1). 1–21. DOI: [10.1017/S1366728901000116](https://doi.org/10.1017/S1366728901000116).
- Myers-Scotton, Carol. 1993. *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon.
- Nazzi, Thierry, Josiane Bertoncini & Jacques Mehler. 1998. Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance* 24(3). 756–766. DOI: [10.1037/0096-1523.24.3.756](https://doi.org/10.1037/0096-1523.24.3.756).
- Otheguy, Ricardo, Ofelia García & Wallis Reid. 2015. Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review* 6(3). 281–307. DOI: [10.1515/applirev-2015-0014](https://doi.org/10.1515/applirev-2015-0014).
- Paradis, Johanne & Fred Genesee. 1996. Syntactic acquisition in bilingual children: Autonomous or interdependent? *Studies in Second Language Acquisition* 18(1). 1–25. DOI: [10.1017/S02722263100014662](https://doi.org/10.1017/S02722263100014662).
- Paradis, Johanne & Samuel Navarro. 2003. Subject realization and crosslinguistic interference in the bilingual acquisition of Spanish and English: What is the role of the input? *Journal of Child Language* 30(2). 371–393. DOI: [10.1017/S0305000903005609](https://doi.org/10.1017/S0305000903005609).
- Paradis, Johanne, Elena Nicoladis & Fred Genesee. 2000. Early emergence of structural constraints on code-mixing: Evidence from French–English bilingual children. *Bilingualism: Language and Cognition* 3(3). 245–261. DOI: [10.1017/S1366728900000365](https://doi.org/10.1017/S1366728900000365).
- Park-Johnson, Sunny K. 2017. Crosslinguistic influence of wh-in-situ questions by Korean-English bilingual children. *International Journal of Bilingualism* 21(4). 419–432. DOI: [10.1177/1367006916629224](https://doi.org/10.1177/1367006916629224).
- Peter, Michelle, Franklin Chang, Julian M. Pine, Ryan Blything & Caroline F. Rowland. 2015. When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language* 81. 1–15. DOI: [10.1016/j.jml.2014.12.002](https://doi.org/10.1016/j.jml.2014.12.002).
- Petersen, Jennifer. 1988. Word-internal code-switching constraints in a bilingual child's grammar. *Linguistics* 26(3). 479–494. DOI: [10.1515/ling.1988.26.3.479](https://doi.org/10.1515/ling.1988.26.3.479).

Marilyn M. Vihman & Virve-Anneli Vihman

- Phillips, Shannon & Margaret Deuchar. 2022. The role of the input in the acquisition of code-switching. In Unn Røyneland & Robert Blackwood (eds.), *Multilingualism across the lifespan* (Routledge Critical Studies in Multilingualism), 56–79. New York: Routledge. DOI: [10.4324/9781003125815-5](https://doi.org/10.4324/9781003125815-5).
- Polka, Linda, Adriel John Orena, Megha Sundara & Jennifer Worrall. 2017. Segmenting words from fluent speech during infancy – challenges and opportunities in a bilingual context. *Developmental Science* 20(1). e12419. DOI: [10.1111/desc.12419](https://doi.org/10.1111/desc.12419).
- Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics* 18(7/8). 581–618. DOI: [10.1515/ling.1980.18.7-8.581](https://doi.org/10.1515/ling.1980.18.7-8.581).
- Qi, Ruying & Bruno Di Biase. 2020. The influence of the environmental language (L_e) in Mandarin-English bilingual development: The case of transfer in wh-questions. *International Journal of Bilingualism* 24(4). 691–714.
- Ramus, Franck. 2002. Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition* 2(1). 85–115. DOI: [10.1075/arla.2.05ram](https://doi.org/10.1075/arla.2.05ram).
- Rodina, Yulia & Marit Westergaard. 2013. The acquisition of gender and declension class in a non-transparent system: Monolinguals and bilinguals. *Studia Linguistica* 67(1). 47–67. DOI: [10.1111/stul.12012](https://doi.org/10.1111/stul.12012).
- Ronjat, Jules. 1913. *Le développement du langage observé chez un enfant bilingue*. Paris: Champion.
- Rothman, Jason, Fatih Bayram, Vincent DeLuca, Grazia Di Pisa, Jon Andoni Duñabeitia, Khadij Gharibi, Jiuzhou Hao, Nadine Kolb, Maki Kubota, Tanja Kupisch, Tim Laméris, Alicia Luque, Brechje van Osch, Sergio Miguel Pereira Soares, Yanina Prystauka, Deniz Tat, Aleksandra Tomić, Toms Voits & Stefanie Wulff. 2023. Monolingual comparative normativity in bilingualism research is out of “control”: Arguments and alternatives. *Applied Psycholinguistics* 44(3). 316–329. DOI: [10.1017/S0142716422000315](https://doi.org/10.1017/S0142716422000315).
- Rowland, Caroline F., Franklin Chang, Ben Ambridge, Julian M. Pine & Elena V. M. Lieven. 2012. The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition* 125(1). 49–63. DOI: [10.1016/j.cognition.2012.06.008](https://doi.org/10.1016/j.cognition.2012.06.008).
- Sağın-Şimşek, Çiğdem & Elena Antonova-Ünlü. 2021. At the syntax-pragmatics interface: Acquisition of Turkish word order by Turkish-English, Turkish-German and Turkish-Russian bilingual children. *Dilbilim Araştırmaları Dergisi* 32. 71–87. DOI: [10.18492/dad.764149](https://doi.org/10.18492/dad.764149).

6 Linguistic interaction in early bilingual development

- Schmitz, Katrin, Marisa Patuto & Natascha Müller. 2012. The null-subject parameter at the interface between syntax and pragmatics: Evidence from bilingual German–Italian, German–French and Italian–French children. *First Language* 32(1–2). 205–238. DOI: [10.1177/0142723711403880](https://doi.org/10.1177/0142723711403880).
- Serratrice, Ludovica. 2005. Language mixing and learning strategy. *International Journal of Bilingualism* 9(2). 159–177. DOI: [10.1177/13670069050090020301](https://doi.org/10.1177/13670069050090020301).
- Serratrice, Ludovica. 2013. Cross-linguistic influence in bilingual development: Determinants and mechanisms. *Linguistic Approaches to Bilingualism* 3(1). 3–25. DOI: [10.1075/lab.3.1.01ser](https://doi.org/10.1075/lab.3.1.01ser).
- Serratrice, Ludovica, Antonella Sorace & Sandra Paoli. 2004. Crosslinguistic influence at the syntax–pragmatics interface: Subjects and objects in English–Italian bilingual and monolingual acquisition. *Bilingualism: Language and Cognition* 7(3). 183–205. DOI: [10.1017/S1366728904001610](https://doi.org/10.1017/S1366728904001610).
- Silva-Corvalán, Carmen & Simona Montanari. 2008. The acquisition of *ser*, *estar* (and *be*) by a Spanish–English bilingual child: The early stages. *Bilingualism: Language and Cognition* 11(3). 341–360. DOI: [10.1017/S136672890800357X](https://doi.org/10.1017/S136672890800357X).
- Sinka, Indra & Christina Schelleter. 1998. Morphosyntactic Development in Bilingual Children. *International Journal of Bilingualism* 2(3). 301–326. DOI: [10.1177/136700699800200303](https://doi.org/10.1177/136700699800200303).
- Soriente, Antonia. 2007. Cross-linguistic and cognitive structures in the acquisition of WH-questions in an Indonesian-Italian bilingual child. In Istvan Kecskes & Liliana Albertazzi (eds.), *Cognitive aspects of bilingualism*, 325–362. Dordrecht: Springer Netherlands. DOI: [10.1007/978-1-4020-5935-3_11](https://doi.org/10.1007/978-1-4020-5935-3_11).
- Soriente, Antonia. 2014. Language development of bilingual children The acquisition of tense and aspect in an Italian-Indonesian child: A case study. *Wacana, Journal of the Humanities of Indonesia* 15(1). 131. DOI: [10.17510/wjhi.v15i1.108](https://doi.org/10.17510/wjhi.v15i1.108).
- Travis, Catherine E., Rena Torres Cacoullos & Evan Kidd. 2017. Cross-language priming: A view from bilingual speech. *Bilingualism: Language and Cognition* 20(2). 283–298. DOI: [10.1017/S1366728915000127](https://doi.org/10.1017/S1366728915000127).
- Tulloch, Michelle K. & Erika Hoff. 2023. Filling lexical gaps and more: Code-switching for the power of expression by young bilinguals. *Journal of Child Language* 50(4). 981–1004. DOI: [10.1017/S0305000922000307](https://doi.org/10.1017/S0305000922000307).
- Unsworth, Sharon. 2003. Testing Hulk & Müller (2000) on crosslinguistic influence: Root Infinitives in a bilingual German/English child. *Bilingualism: Language and Cognition* 6(2). 143–158. DOI: [10.1017/S1366728903001093](https://doi.org/10.1017/S1366728903001093).
- Valdés Kroff, Jorge R. & Paola E. Dussias. 2023. Production, processing, and prediction in bilingual codeswitching. In Kara D. Federmeier & Jessica L. Montag (eds.), *Speaking, writing and communicating* (Psychology of learning and motivation 78), 195–237. Cambridge: Elsevier. DOI: [10.1016/bs.plm.2023.02.004](https://doi.org/10.1016/bs.plm.2023.02.004).

Marilyn M. Vihman & Virve-Anneli Vihman

- van Osch, Brechje, Maria del Carmen Parafita Couto, Ivo Boers & Bo Sterken. 2023. Adjective position in the code-switched speech of Spanish and Papiamentu heritage speakers in the Netherlands: Individual differences and methodological considerations. *Frontiers in Psychology* 14. DOI: [10.3389/fpsyg.2023.1136023](https://doi.org/10.3389/fpsyg.2023.1136023).
- Van Dijk, Chantal, Elise Van Wonderen, Elly Koutamanis, Gerrit Jan Kootstra, Ton Dijkstra & Sharon Unsworth. 2022. Cross-linguistic influence in simultaneous and early sequential bilingual children: A meta-analysis. *Journal of Child Language* 49(5). 897–929. DOI: [10.1017/S0305000921000337](https://doi.org/10.1017/S0305000921000337).
- Vasilyeva, Marina, Heidi Waterfall, Perla B. Gámez, Ligia E. Gómez, Edmond Bowers & Priya Shimpi. 2010. Cross-linguistic syntactic priming in bilingual children. *Journal of Child Language* 37(5). 1047–1064. DOI: [10.1017 / S0305000909990213](https://doi.org/10.1017/S0305000909990213).
- Vihman, Marilyn. 1985. Language differentiation by the bilingual infant. *Journal of Child Language* 12. 297–324.
- Vihman, Marilyn. 1998. A developmental perspective on codeswitching: Conversations between a pair of bilingual siblings. *International Journal of Bilingualism* 2(1). 45–84. DOI: [10.1177/136700699800200103](https://doi.org/10.1177/136700699800200103).
- Vihman, Marilyn. 2014. *Phonological development: The first two years*. 2nd edn. Malden: Wiley-Blackwell.
- Vihman, Marilyn May. 1982. Formulas in first and second language acquisition. In Loraine K. Obler & Lise Menn (eds.), *Exceptional language and linguistics*, 261–284. New York: Academic Press.
- Vihman, Virve-Anneli. 2016. Code-switching in emergent grammars: Verb marking in bilingual children’s speech. *Philologia Estonica Tallinnensis* 1. 154–172. DOI: [10.22601/PET.2016.01.10](https://doi.org/10.22601/PET.2016.01.10).
- Vihman, Virve-Anneli. 2018. Language interaction in emergent grammars: Morphology and word order in bilingual children’s code-switching. *Languages* 3(4). 1–24. DOI: [10.3390/languages3040040](https://doi.org/10.3390/languages3040040).
- Villa-García, Julio & Imanol Suárez-Palma. 2016. Early null and overt subjects in the Spanish of simultaneous English-Spanish bilinguals and Crosslinguistic Influence. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics* 29(2). 350–395. DOI: [10.1075/resla.29.2.01vil](https://doi.org/10.1075/resla.29.2.01vil).
- Volterra, Virginia & Traute Taeschner. 1978. The acquisition and development of language by bilingual children. *Journal of Child Language* 5(2). 311–326. DOI: [10.1017/S0305000900007492](https://doi.org/10.1017/S0305000900007492).
- Wei, Li. 2018. Translanguaging as a practical theory of language. *Applied Linguistics* 39(1). 9–30. DOI: [10.1093/applin/amx039](https://doi.org/10.1093/applin/amx039).

6 *Linguistic interaction in early bilingual development*

- Wolleb, Anna, Antonella Sorace & Marit Westergaard. 2018. Exploring the role of cognitive control in syntactic processing: Evidence from cross-language priming in bilingual children. *Linguistic Approaches to Bilingualism* 8(5). 606–636. DOI: [10.1075/lab.17002.wol](https://doi.org/10.1075/lab.17002.wol).
- Yip, Virginia & Stephen Matthews. 2000. Syntactic transfer in a Cantonese-English bilingual child. *Bilingualism: Language and Cognition* 3(3). 193–208. DOI: [10.1017/S136672890000033X](https://doi.org/10.1017/S136672890000033X).
- Yip, Virginia & Stephen Matthews. 2016. Code-mixing and mixed verbs in Cantonese-English bilingual children: Input and innovation. *Languages* 1(1). 4. DOI: [10.3390/languages1010004](https://doi.org/10.3390/languages1010004).
- Yip, Virginia & Stephen Matthews. 2022. Language diversity and bilingual first language acquisition: A commentary on Kidd and Garcia (2022). *First Language* 42(6). 832–836. DOI: [10.1177/01427237221097581](https://doi.org/10.1177/01427237221097581).

Chapter 7

Supporting bilingual development through code-switching: Dual language use of a grandmother and an emergent bilingual child

Suzanne Quay

International Christian University, Tokyo

Children exposed predominantly to a heritage language at home usually find it difficult to fit into environments where a societal language is used. In this case study, a Cantonese-English bilingual grandmother in western Canada, concerned about her grandson's lack of interaction at preschool due to limited English ability, code-switched to provide more input in the societal language to balance her grandson's bilingualism. This investigation focuses on how a code-switching caregiver can support an emergent bilingual child's weaker language and how the child uses his entire linguistic repertoire to accommodate and communicate in response.

Video-recordings, made weekly at the grandmother's home when the child was between 2;9–4;11 of age, were used for quantitative analyses of language choice, and for qualitative analyses of code-switching and discourse patterns. While increasing his exposure to English, the grandmother used Cantonese to scaffold and negotiate meaning with her grandson. The child was able to accommodate his grandmother's use of English by using congruent lexicalization and convergence as part of his code-switching practices not seen in his grandmother's data: (7) to convey information in 'mixed' complex/compound sentences that he could not produce in one language alone, and (8) to form complete utterances in his weaker one by using the morphosyntactic rules of his stronger language. The fluidity and hybridity of the conversations between the child and his grandmother show how such translingual practices can provide a linguistic bridge for emergent bilingualism.

Suzanne Quay

1 Introduction

Studies show that Asian families in English-speaking countries like Australia, Canada and the US believe that proficiency in the *heritage* or *home language* (HL for either terminology to denote *non-societal* languages, but see Eisenchlas & Schalley (2020), for conceptual distinctions) give children a sense of cultural identity (Liang & Shin 2021, Shen & Jiang 2021), can be valuable for mobility and job prospects in the future (Mu 2014), and is important for communication with grandparents and relatives (Park & Sarkar 2007, Tran et al. 2022). To maintain HLs, parents must make family language policies (FLP) about language use within the home and among family members. King et al. (2008: 907) consider this necessary not only for shaping “children’s developmental trajectories” to connect with “formal school success” but also for “the maintenance and future status of minority languages” (for a review of FLP research, see Lanza & Lomeu Gomes 2020). Maintaining HLs is indeed particularly difficult when children know that their “parents are highly competent users of English” as Little (2020b: 210) discovered in her survey of 212 bilingual families in the UK.

1.1 Difficulties in maintaining home/heritage languages

Many bilingual families find it challenging to preserve HLs in English-dominant environments when their efforts are met with disapproval due to linguistic prejudice in society (Schroedler et al. 2024) or to economic and practical concerns inherent in majority-minority language issues (Spolsky 2012). The *emergent* bilingualism (defined in García et al. 2008) of HL-dominant children is not appreciated when they start school without adequate proficiency in the societal language. Even in bilingual preschools in the US and Canada, children tend to drop their HL soon after learning English as they “can tell by the way people interact with them that the only language that counts for much is English” (Wong Fillmore 1991: 341). A task-based study of 58 Greek-heritage children in Canada and the US found that third generation children had significantly lower accuracy in HL tasks than second and mixed-generation (having first- and second-generation parents) children (Chondrogianni & Daskalaki 2023). This is corroborated by census data in the US (Alba et al. 2002) and Canada (Schott et al. 2021) that show that immigrant families are assimilating into the dominant culture and language by the second generation with total HL loss by the third. Children are more likely to be actively bilingual in Canada if they have a prestigious HL (e.g., French), if their parents are highly educated and can afford French-English educational programs, and if supplementary afterschool and weekend programs exist for their

7 Supporting bilingual development through code-switching

HL (Slavkov 2017). Since most children with HLs are enrolled in English-medium education, they are likely to shift to English due to the “marginalisation of heritage languages in public and political discourse” (Meddegama 2020: 644, and Alba et al. 2002).

Parents face various dilemmas when they make FLP decisions. If they have only been using the HL from birth, when should they introduce the societal language? And how much exposure should they give to each language? In discussing *harmonious* bilingualism (positive subjective experiences in bilingual settings), De Houwer (2020: 68) emphasizes that “societal language proficiency is important for all children” and that “bilingual children who have developed good levels of proficiency in the societal language upon (pre)school entry have an advantage over emergent bilingual children in terms of well-being” (see De Houwer 2020, for supporting literature). Schroedler et al. (2024) describe the disadvantages faced by emergent bilinguals in Germany with less prestigious HLs (e.g., Turkish) than English, French, Spanish and Italian. Severe discrimination from teachers (and unkind teasing from classmates) are reported starkly in personal accounts of primary school experiences.

Without compromising HL development, we need to determine the best way for young children to acquire the societal language before formal education. This study explores the action taken by a grandmother who was worried about her grandson’s well-being when he could not communicate in his preschool’s English environment because he was dominant in the HL established by his parents’ FLP. The grandmother increased her grandson’s exposure to the societal language by code-switching in conversations with him. Her role in helping the child to become stronger in the societal language before the start of formal schooling will be investigated.

1.2 The role of grandparents in language development

Many fervent calls exist in the *Handbook of Home Language Maintenance and Development* (Schalley & Eisenclas 2020) and *The Cambridge Handbook of Childhood Multilingualism* (Stavans & Jessner 2022) for more research involving different generations and various family configurations. This call has been answered somewhat in studies that address the impact of grandparents on HL transmission (Meddegama 2020, Spolsky 2012). In Japan where the pressure for homogeneity still prevails, grandparents’ attitudes towards foreign daughters-in-law and their languages can support or jeopardize HL transmission. The monolingual grandmother in Nakamura & Quay (2012) participated willingly in supporting her grandson’s bilingual development in a one-person-one-language setting where

Suzanne Quay

she provided Japanese input while her daughter-in-law provided English, a prestigious language in Japan. Unfortunately, Japanese grandparents living in the same household who disapproved of their Thai and Filipino daughters-in-law inhibited the use of Thai and Tagalog, respectively, in the home so that three mixed ethnic children were raised only with Japanese. Nakamura (2020) revealed that the only two mixed ethnic participants raised bilingually had Filipino mothers who were estranged from their Japanese paternal grandparents and did not live with them. Using the US Census 2000 Supplementary Survey, Ishizawa (2004) concluded that non-English-speaking grandparents in three-generation households influenced grandchildren's HL maintenance. She found that the presence of grandmothers (more than grandfathers) and paternal (more than maternal) grandparents had a stronger effect on children's minority language use. In trilingual families in Germany and England, Braun (2012) discovered that monolingual grandparents were more successful than bilingual ones in helping to pass on native languages and cultures to grandchildren. Bilingual grandparents living in the same community tended to disregard the FLP of parents and used the societal language they deemed to be more valuable, which frustrated the parents and deprived the grandchildren of the necessary input to maintain the minority language. In a family with a strict Gaelic FLP in Scotland, the paternal grandmother, with the children's mother, spoke Gaelic to her two grandchildren, but the father, with his siblings, undermined these efforts by using English extensively and Gaelic only as a language of discipline (Smith-Christmas 2014).

Grandparents can indeed be important minority language resources as the main or secondary providers of HL input. Silva-Corvalán (2014) became the sole Spanish interlocutor for her two grandsons in their first six years of life in the US when the two boys' father stopped using Spanish with them after age 3;6. Despite the family's strong desire to raise bilingual children, the father (as in Smith-Christmas 2014) started using more and more English after that age because all other family members were English speakers. The two boys ended up hearing Spanish 20%–35% of the time predominantly from their grandmother until age 6;0. Silva-Corvalán (2014: 348) concluded that lower exposure to Spanish had "a direct consequence on the level of proficiency attained in different grammatical domains" with the older grandson producing more grammatically correct Spanish utterances than the younger one who had less Spanish exposure. In Montanari (2009), two sets of grandparents helped a trilingual child's parents to support two HLs despite the ubiquity of English in the US; the maternal grandparents supplemented the child's exposure to Tagalog from her mother while the paternal grandparents provided additional Spanish exposure with the father. In Ruby (2012: 81), a Bangladeshi grandmother in London code-switched

7 Supporting bilingual development through code-switching

into Bangla to teach the HL informally while still using English to meet her third-generation English-dominant six-year-old granddaughter “halfway”. Even before the global pandemic, digital intergenerational interactions have become the focus of language practices in multilingual families. Although physically absent, grandparents have provided their grandchildren with informal HL learning through digital communication (e.g., texting, using social media platforms, etc.) (see [Lanza & Lexander 2019](#) and [Palviainen 2020](#) for an overview of such communication).

Grandparents have also helped grandchildren in language and literacy practices in the transition from home to school. A survey of grandparents in 20 families with children aged 3–6 in East London described how *scaffolding* by grandparents while story reading, writing in two languages or playing word games helped grandchildren in school learning ([Kenner et al. 2007](#)). This transfer of knowledge went both ways as grandchildren also taught grandparents how to use computers and information technology. In fact, grandparents and grandchildren “used their different capabilities to create shared understandings, leading to new forms of linguistic and cultural learning” ([Kenner et al. 2007](#): 237) reminiscent of *translanguaging* activities suggested for educational contexts ([García 2009](#)). Also in London, a Bengali grandmother told her grandchildren European traditional tales (e.g., Snow White) in Bengali, which helped them understand the same stories in English better at school ([Gregory et al. 2007](#)). In Montreal and Singapore, Chinese grandmothers modelled and implicitly demonstrated literacy use for a variety of purposes (from wordplay involving writing and guessing riddles to structured literacy practice using workbooks) that encouraged their grandchildren to become literate in multiple languages ([Curd-Christiansen 2013](#)).

Children may become motivated to maintain HLs because of emotional bonds with their grandparents. [Melo-Pfeifer \(2015: 40\)](#) found that amongst a younger generation of Portuguese immigrants in Germany, “affective and emotional roles seem to be particularly linked to grandparents, which becomes evident when children evoke Portugal and visiting Portugal or when they mention their motivations to keep on learning Portuguese.” Conversely, harmonious bilingualism does not occur when critical grandparents trigger anxiety across generations. Turkish grandparents lowered the third generation’s self-esteem and demotivated them from speaking Turkish in the Netherlands when they strongly criticized their third-generation Turkish-Dutch bilingual grandchildren’s poor use of the HL ([Sevinç 2020](#)).

Suzanne Quay

2 Dual language input in bilingual development

Whether children become active bilinguals depends on the quantity and quality of input they receive in each language (Schalley & Eisenchlas 2022) and the timing of language exposure (which is of greater importance for minority than for majority language development; e.g., Hoff & Core 2013, Montrul 2022). The impact of language input on development also depends on having a diversity of language contact through exposure to different family members, others in the community, electronic media and printed sources (see Quay & Montanari 2016 and Quay & Chevalier 2019 for further discussion; for an overview of the use of social media and technology, see Little 2020a). Any variation in the quantity and quality of input in each language can affect the rate at which each language is learned (e.g., Cantone 2022) and may also result in language change across generations. Complex interactions occur between child internal factors like their language preference and personality, and child external factors such as the quantity and quality of the input, particularly in how caregivers respond to language mixing (Chevalier 2013, Quay 2012).

However, young children are not the only ones who mix languages. Caregivers can choose to provide bilingual input to young children by alternating two languages through code-switching *inter-sententially* (between sentences) or *intra-sententially* (within a sentence). Code-switching in adults is a typical way of speaking among bilinguals that reflect bilingual proficiency rather than deficiency (Bullock 2009, MacSwan 2016). Researchers often use the term *code-switching* interchangeably with *code-mixing* when referring to young learners' speech, particularly in the early stages of bilingual development (Quin et al. 2016), with some like Meisel (1994) attempting to differentiate the two terms according to whether children have acquired proficiency in the two languages (see also Nicoladis 2013 who point out that some researchers use code-mixing exclusively to indicate that children's productions may be less deliberate than the code-switching of adults). In this chapter, the two terms, code-switching and code-mixing, are used interchangeably without implying any language deficiency even in young bilingual learners.

Indeed, when bilingual children first start speaking, they generally code-mix more to fill lexical gaps when trying to speak their weaker language than their stronger one (e.g., Deuchar & Quay 1999), but they do so while accommodating to the input in their environments (e.g., Lanza 1997). Input conditions thus play a crucial role on the rates and directionality of code-mixing (see Deuchar 2022 for a more comprehensive review). For example, simultaneous French-English bilingual children under age 3 in Montreal were found to be able to adjust their rate

7 Supporting bilingual development through code-switching

of code-mixing (and even repair breakdowns in communication due to language choice) in response to the code-mixing rate of interlocutors known (parents) and unknown (strangers) to them, thus demonstrating differentiation of the two languages even when this required using a weaker language (Comeau et al. 2003, 2007). In a Hong Kong study of Cantonese-English bilingual children aged 1–4, Yip & Matthews (2016) report that two children from one-parent-two-language environments code-mixed twice as often as seven children from one-parent-one-language environments.

Although it is commonly believed that acquiring two languages in a code-switching environment will lead to language-related deficit, MacSwan et al. (2019) provide evidence to the contrary. They found that code-switching (even intrasententially) in the home had no negative consequences for bilingual language acquisition when they compared the Spanish and English test results from two groups of 11-to-12-year-old bilingual children who had started learning English only in kindergarten. The 38 from non-code-switching homes and 42 from code-switching homes performed similarly in all the assessments in each language by Grade 6 whether they were raised in a code-switching home environment or not (note, however, that the authors do caution that that they did not have observational data to confirm the frequency of code-switching practices reported by the children’s parents on a survey about their home language environment). Indeed, code-switching was advantageous in Singapore in giving 5- to 6-year-old children a chance to use English and Mandarin that helped them develop bilingual proficiency particularly in their weaker language (Yow et al. 2018). This advantage had already been found in studies conducted in US public schools in the late 1980s and early 1990s as described by Faltis (2019: 47) about pedagogical code-switching where “students engaged in protracted interaction in one language and then the other. The NCA [New Concurrent Approach] was the first bilingual pedagogy to offer a viable and effective way of promoting both language and content development during an era when any kind of language mixing during bilingual instruction was viewed as harmful to students.” In this approach, the recommendation to use only intersentential code-switching was limited to classroom teachers (and not expected of children) to ensure “quality bilingual instruction and interaction” that “include long chunks of language between two named languages (not language separated by teachers, days, classroom) throughout the day, so that children hear and respond to ideas and content in both languages, with opportunities to stretch their discourse abilities” (Faltis 2019: 49–50). Although beyond the scope of this paper, it is necessary to mention the similarities between *pedagogical code-switching* and the trending term *pedagogical translanguaging*. Both terms refer to the strategic use of

Suzanne Quay

two or more languages in classroom instruction to activate students' resources from their whole linguistic repertoire (Cenoz & Gorter 2022). Note, however, that the meaning of the term *translanguaging* has changed over time (see, e.g., Balam 2021). Gort (2019: 165) points out, however, that code-switching gives us a “nuanced window into the sophisticated, rule-governed linguistic behavior of bilinguals that is characterized by grammatical systematicity and pragmatic coherence,” while translanguaging “allows us to understand and conceptualize the broad range of language practices of bilinguals ... without reference to existing notions of grammaticality.” In a comprehensive review of studies of code-switching among bilingual and trilingual children, Treffers-Daller (2022: 193) reports not only that code-switching input does not have a negative impact on children's language development in each language, but that intrasentential code-switching is “not a sign of deficiency in grammar or pragmatics, but rather an illustration of multilinguals' ability to creatively exploit the different resources at their disposal.”

2.1 Code-switching studies: From rules and constraints to typology

That code-switching can be beneficial for bilingual learning and is also a sign of proficiency is further supported by studies over the last half century showing it is governed by rules and constraints. These studies have directed much effort to locating regular patterns that may be found in instances of *intra*-sentential code-switching. MacSwan (2019: 13, Table 1.2) provides a historical summary of the study of code-switching as linguistic structure. For example, phrase-structure oriented proposals from the 1980s to 2010s started with the *equivalence constraint* predicting that codeswitches tend to occur at points where the combination of elements from the two languages do not violate a syntactic rule of either language and the *free morpheme constraint* predicting a “switch may not occur between a bound morpheme and a lexical form unless the latter has been phonologically integrated into the language of the bound morpheme” (Sankoff & Poplack 1981: 5). That era ends with the Matrix Language Frame model proposed by Myers-Scotton (1993) whereby the base or matrix language (ML) sets the frame to receive words or phrases from the embedded language (EL) with three categories: (1) ML clauses or phrases follow ML syntactical rules, (2) EL phrases follow EL syntactical rules, and (3) EL elements are integrated into the syntax and often into the morphology of the ML when constituents contain both EL and ML elements. This is an asymmetric approach that assumes that one language provides the grammatical rules for a given clause while items from the other language is inserted into the structure of the first language. This was exemplified in examples

7 Supporting bilingual development through code-switching

provided by Yip & Matthews (2007) from Cantonese-dominant bilingual children, aged 2;6–4, who formed relative clauses in English using the Cantonese pattern. Such crosslinguistic influences are more likely when the grammatical constraints or properties of each language differ but overlap (e.g., Hulk & Müller 2000). Yip & Matthews (2000) conclude that the pervasiveness of transfer implies a high degree of interaction between the two distinct and separate linguistic systems that are simultaneously developing in the bilingual mind. Such examples reflect the differentiated but dynamic nature of bilingual development (Genesee 2022).

In showing how the code-switching literature and field has evolved and self-corrected over time, MacSwan (2019: 20) also provides a concise review of the current (2000s to present) Feature Oriented Proposals that are ‘constraint-free’ approaches that do “not posit rules or conditions that are specific to code-switching but rather allows the grammaticality facts and patterns to fall out of independently motivated principles of grammar.” This is a symmetrical approach also known as minimalism that does not assume priority for any one language but allows lexical items to come from either language and for their features to determine the possibility of their combination; that is, grammaticality in code-switching emerges from the interaction of the bilingual’s grammar rather than from subconscious rules governing language mixing (see Deuchar 2013 for an illustration of this approach with concrete examples). It is beyond the scope of this chapter to discuss bilingual grammar further (different proposals exist, e.g., López 2020, MacSwan 2022).

Instead of formulating rules for intrasentential code-switching and then checking with data to see if the rules apply as described for the above approaches, Treffers-Daller (2022) found it more fruitful to review the different types of code-switching that multilingual children use according to the typological approach outlined for adults by Muysken (1997, 2000). Methodologically speaking, this approach allows researchers to use naturalistic data without making assumptions about the data in advance, except that they will follow one of the four (three in the 1997 publication) patterns outlined by Muysken (2000): (7) *insertion* or the use of a content word (like a noun) or multi-word chunks in a grammatical frame comprising mainly words in the other language; (8) *alternation* where one language switches to another involving both grammar and lexicon not only between clauses in a single turn but also between conversational turns; (9) *congruent lexicalization* when the two languages share a structure filled with elements from either language; and (10) *backflagging* when speakers use HL discourse markers in their L2 utterances (see Treffers-Daller 2022 for examples of each pattern). This typological approach makes it easier to relate different code-switching patterns to sociolinguistic factors (like the status of the languages in the communities

Suzanne Quay

or the children's proficiency). Treffers-Daller (2022) found that among the four patterns, congruent lexicalization has not been frequently discussed in the literature on children's code-switching but presents two examples from Gawlitzek-Maiwald & Tracy (1996) who use a bilingual bootstrapping metaphor to explain why this pattern occurred in their German-English data. Congruent lexicalization as part of language mixing is believed to bridge structural as well as lexical gaps wherein the child's stronger language provides grammatical structures not yet available in the weaker language. The term *bilingual bootstrapping* is also used by the authors to avoid the negative connotations associated with crosslinguistic transfer.

2.2 The current investigation

The current case study is different from most FLP studies that use interviews (Shen & Jiang 2021), focus groups (Tran et al. 2022) or large survey data (Little 2020b, Mu 2014) to investigate HL maintenance. Instead, longitudinal weekly recordings of a preschooler and his grandmother allow for micro-level analysis and data-driven insights into the transition from home to school language.

In this study, the parents wanted to expose their son to as much Cantonese as possible before formal education to give him a head-start in the minority language. They knew from experience that English would eventually become his stronger language in Canada and assumed the grandparents on both sides would help in this endeavor. However, as Spolsky (2012: 5) indicates, key participants in FLP other than parents "will have different language practices, different beliefs about the values of the varieties that make up the sociolinguistic ecology of the community, and each may attempt to manage or influence the language practices and beliefs of others." Because FLP is not static, "it is critical to track changes of FLP over time [...] to identify factors that have influence on children's language behaviour and social development and during what period this influence is most important" (Curdtt-Christiansen & Sun 2022: 272). The paternal grandmother stopped following the parents' FLP when her grandson started preschool. She strongly believed he needed English to interact with others in his schooling, which made FLP, in her opinion, secondary to successful educational integration. This study highlights the challenges between each family members' ideologies and personal multilingual experiences (see Purkarthofer 2020 for more on intergenerational challenges in language practices).

The purpose of this study is to investigate bilingual input to a child from a code-switching grandmother to understand how her speech style affected her grandson's emergent bilingualism during the preschool period:

7 Supporting bilingual development through code-switching

- (1) Are there quantitative and qualitative differences in dual-language use by the grandmother and the grandson?
- (2) How does the grandmother use bilingual input to support her grandson's weaker language?
- (3) What developmental features can be gleaned from the child's linguistic repertoire?

The first research question addresses code-switching as a common phenomenon among bilingual speakers and as part of FLP and language socialization. The next two consider its use and effects on the preschooler's development of his weaker societal language. Does the grandson's linguistic repertoire reflect the code-switching practices of his grandmother or reveal any innovations?

3 Methodology

3.1 Case study background

The family lived in western Canada in the same metro area as two sets of grandparents. The mother estimated that her second-generation firstborn child, Ryan (RYA), was exposed from birth until age five to 90% Cantonese from her, 60% from his father and 99% from his maternal grandparents who lived close by. From age one, his paternal grandfather addressed him in Mandarin for short periods when they were alone together. The child was dominant in Cantonese (with minimal English) before his paternal grandmother (PGM) started taking care of him regularly when he began attending a drop-in preschool where he received 4–5 hours of English input each week. Unlike the Chinese-dominant maternal grandparents, the paternal grandparents (and parents) had all gone through English-medium education.

PGM, a regular caregiver on the days when RYA attended preschool in the mornings, noticed how quiet her grandson was in the preschool environment when surrounded by other children and their accompanying English-speaking caregivers. A 26-minute preschool recording at age 3;0.26 (year;month.day) confirms her observations. When the teacher questioned him and three other children about a picture book, the other children supplied one-word English labels for objects in the story while he only pointed at them. He did repeat one-word utterances twelve times when urged by PGM and the teacher to do so but had only one spontaneous utterance, "upside down," in that recording. Thus, the grandmother felt compelled to provide more input in English although she previously

Suzanne Quay

used mainly Cantonese as his parents requested. The paternal grandparents (unlike the maternal ones) spoke Cantonese and English to each other. Both were dominant in English due to their education and work before retirement. The grandfather – multilingual in English and seven Chinese dialects (in acquisition order, Fukian, Teochew, Cantonese, Mandarin, Toisan, Hakka and Hainanese) – only spoke Cantonese and English at home. To maintain his Mandarin, he watched Mandarin television programs and used it at home only with his grandson.

3.2 Data collection and analyses

With parental and extended-family consent, PGM was asked to collect data weekly between ages 2;9.1–4;10.14 using a video camera set up on a tripod at her home. Fifty-four sessions were selected from this two-year period and transcribed in the CHAT format of CHILDES (Cantonese utterances were romanized without transcription of lexical tones). Sixteen transcripts (totalling 460 minutes and averaging 28.75 minutes/session) when PGM and RYA were alone were selected for an analysis of their language use. In these sessions, PGM read books with RYA, played his favorite board game, Snakes & Ladders, or card games, and used flashcards to teach the alphabet, spelling, and proper names of his family members. They also did Math workbooks together.

All utterances were coded for language and extracted from the transcripts by using the KWAL (Key Word and Line) command of the CLAN (Computerized Language Analysis) program. The rate of intersentential and intrasentential code-switching was determined by dividing the number of utterances coded as ENG (English), CHI (Cantonese) and MIX (both English and Cantonese) for each interlocutor with their total utterances per session. MIX utterances were categorized further as:

- (4) One-word insertions: Predominantly English (ENGchi) or Cantonese (CHIeng) with one lexical item from the other language as in “*ikoh* [this] fish is nice” or “*ngo mou sai* [I don’t have any more] ketchup”.
- (5) Two-word mixes: Two-word utterances with one element from each language (TWOmix) as in “more *tong* [soup]” or “*itou* [here] wrong”.
- (6) Multi-word mix utterances: English-Cantonese utterances beyond one-word insertions (ENG-CHI examples shown later).

Quantitative analyses were conducted to determine whether PGM and RYA’s language choices matched. RYA’s utterances were also analyzed qualitatively to

7 Supporting bilingual development through code-switching

examine the effect of PGM's code-switching on his weaker language, English. The child's English usage in ENG-CHI and ENG utterances was also translated into Cantonese to examine for congruent lexicalization and convergence.

3.3 Results

3.4 The grandmother and child's discourse patterns

Table 1 provides an overview of the quantity of utterances produced by PGM and RYA from 2;9–4;10 in 16 sessions. Excluding at age 4;3.17 (when he was talking to himself while playing with a train), the child produced, on average, 33% of the number of PGM's utterances in 15 sessions. Without session 11, he produced only an average of 28% of PGM's total utterances up to age 4;5.19. Three months later at age 4;8.19, he produced 66% of PGM's total ($n = 775$) with his highest proportion of 70% produced in the final session 5 months later. Thus, the child started contributing much more to conversations between 19–22 months after his grandmother started exposing him to more English.

Because the child spoke less than the grandmother, calculating in percentage the proportion of their utterances in each language was needed to examine whether their language choices match. Figure 1 shows PGM used more English from 50% of all her utterances in the first session to the last one at 92%. Her language choice had a direct distributional effect on her grandson's production (the blue and red lines show RYA consistently produced more English than Cantonese when alternating languages like PGM). During those two years, when PGM provided on average 65.6% ($n = 7344$) English input, RYA responded with 63.7% ($n = 2345$) English, so he seems to have been accommodating her even in his weakest language. When she provided on average 23.7% ($n = 2649$) Cantonese input, he produced 17% ($n = 627$) Cantonese utterances. He averages more *mixed* utterances at 19% ($n = 709$) than PGM at 11% ($n = 1203$).

Figure 1: Percentage of Cantonese, English and Mixed utterances in Grandmother's input and Ryan's output

3.5 Intrasentential code-switching: Quantity and quality

Both had the lowest number and percentage of mixed utterances roughly matching in percentage for each type of intrasentential switching in Table 2. Their mixed utterances were mostly one-word insertions into multiword utterances in the other language. Adding ENGchi and CHIeng utterances together, one-word

Suzanne Quay

Table 1: The number of utterances and RYA’s proportion of PGM’s utterances in each session

Session	Age	No. of PGM’s utterances	No. of RYA’s utterances	% of RYA’s utterances to PGM
1	2;9.25	971	130	13
2	3;3.25	388	90	23
3	3;4.9	772	310	40
4	3;4.30	950	273	29
5	3;5.27	704	231	33
6	3;6.13	596	154	26
7	3;7.10	960	101	11
8	3;7.24	880	228	26
9	3;8.19	676	279	41
10	3;9.14	489	163	33
11	4;3.17	197	260	132*
12	4;4.10	839	270	32
13	4;5.5	864	171	20
14	4;5.19	744	240	32
15	4;8.19	775	509	66
16	4;10.14	391	272	70
Total		11,196	3681	Av. w/o * 33%

insertions make up 69% of PGM’s mixed utterances (n = 825) and 62% of RYA’s (n = 439). The whole family always used Cantonese kinship terms to show respect for familial relationships and status based on gender and seniority by using specific terms such as *maamaa* and *yeye* for *paternal* grandmother and grandfather, respectively. Thus, RYA had no choice but to use these Cantonese terms of address when speaking in English. They make up the bulk of one-word insertions in predominantly English utterances (ENGchi in Table 2) for both PGM (64%) and RYA (67%) and are not truly intrasentential switches as they behave more like proper names that do not change. They had the same proportion at 8% of two-word mixed utterances such as: “kiss *kumyong*” ‘like this’ (2;9.25); “*pinkoh* ‘which-one’ driver” (3;7.10); “*mou* ‘no’ ink” (4;5.19) from PGM, and “where *dinwa*” ‘phone’ (2;9.25); “*wakje* ‘perhaps’ yes” (4;3.17); “*kum* ‘so’ small” (4;8.19) from RYA. The child did have a higher proportion of multiword mixed utterances at 30%

7 Supporting bilingual development through code-switching

(211/709) than PGM at 23% (269/1194). An example from RYA would be “this *kiew chou* ‘(is) called’ dark blue” (3;5.27) while PGM’s example shows interclausal as well as intrasentential code-switching: “go inside here *mei unwui chengtou teiha lo*” ‘so (you) don’t dirty (the) floor’ (3;3.25).

Table 2: Type of intrasentential code-switching by PGM
(N = 1194) and RYA (N = 709) in 16 sessions

Type of intrasentential CS	PGM			RYA		
	# of utts.	%	Kinship terms	# of utts.	%	Kinship terms
One-CHI word (ENGchi)	222	19	141 (64%)	167	24	112 (67%)
One-ENG word (CHIeng)	603	51		272	38	
Total one-word insertions	825	69		439	62	
TWOmix: two-word utts.	100	8		59	8	
ENG-CHI: multi-word utts.	269	23		211	30	

4 Discussion

4.1 Grandmother’s code-switching practices as support for the weaker language

The emergent bilingual child was pragmatically sensitive to the language choice of his grandmother despite his limited ability in English. From the early recordings onwards, RYA matched PGM’s English input with a similar proportion of English utterances (but see next section for *qualitative* differences in the child’s English output). Figure 1 and Table 2 showed that RYA was sensitive to PGM’s 89% intersentential and 11% intrasentential code alternations. As he got older, she increased her English utterances.

The way PGM used English with Cantonese to support RYA’s English development is reminiscent of studies of classroom code-switching as outlined by Faltis (2019: 49, Table 2.1). PGM provided English vocabulary and encouraged collaborative construction of meaning while she read to him. In Example (7) from the first recording, she supplied translation equivalents and gave additional information in RYA’s stronger language to help him understand the weaker one (more Cantonese was used here than in later sessions). She scaffolds the story for her

Suzanne Quay

grandson by switching from English to Cantonese to provide translations, elaborate on the meaning of English utterances or confirm his comprehension. Her turns are long as she tries to reinforce the main ideas from the story and encourage discussion even though RYA responds (as at preschool) through actions rather than words. PGM's Cantonese utterances are in italics with English glosses between single quotes and English insertions in predominantly Cantonese utterances underlined>.

- (7) PGM is reading from a book to RYA at age 2;9.25

1 *PGM: There is a big cat there. Have you seen a cat? *Nei tung ngo taiha* 'have a look with me'. See *kohko* 'that' cat. *Nei tai keitoh* cat *ah* 'you (can) see so many cat(s)'. What is the cat doing? Huh? What is cat...what is *yeye* 'grandpa' doing? What is *yeye* doing? He is eating. You see this cat. One two. You see the eyes so big. You've got big eyes? How many cats here? Busy, busy kitties. What are they doing?

*RYA: 0.

%act: flips page

2 *PGM: Kitty's sitting. *Huikeh tau kum chai lok hui* '(it) puts its head down like that'. *Hui wan ye hai yappin koh box tou wan ye* 'it's looking for something inside the box'. *Ikoh le* 'what about this'? Kitty peeking. *Chongchongchong* '(It's) peeking'. *Huikeh ngan kumyong chongchong* 'its eyes are peeking like that'. *Timyong chongchong* 'how does (it) peek'?

*RYA: 0.

%act: opens eyes wide

3 *PGM: Oh. And then this kitty *hang lou* 'walks along'. Strolling along. *Ikoh le* 'what about this'? Kitty rolling. Roll. *Kumyong chuen chuen chuen* 'roll roll roll like that'. *Tai hui kumyong chuen chuen chuen hui keh* body 'see how it rolls rolls rolls its body'. Kitty hissing. *Kei auk eh hui ngau* 'how fiercely it bites'. *Huitit nga kei lei ah* 'how sharp are its teeth'. Don't go near a cat. Little kitten. *Tai huikeh nga kei sharp ah haimou* 'its teeth are so sharp aren't they'? *Ngau yan keh haimou* 'it bites people doesn't it'? *Nei yat chengtou hui hui chau ngau neikeh sau kelo nei chimou* 'do you know that if you offend it, it will bite your hand'? *Ikoh* 'this' kitty kissing. Kissing kissing. *Tai hui leongko kiss kiss haimou* 'can you see that the two of them are kissing'? Kiss kiss. *Nah ikoh* 'here this' kitty is running. *Timyong chau ah* 'how does it run'? *Chau chau chau chau kumyong ah* 'does it run run run run like this'?

%sit: PGM pretends to run with downturned index and middle fingers along the sofa

7 Supporting bilingual development through code-switching

This example shows mainly intersentential codeswitches where the utterances are English-only or Cantonese-only (the code-switching pattern recommended for teachers in Faltis 2019). Eight utterances exemplify intrasentential codeswitches with four depicting one-word English insertions (“cat”, “box”, “body” and “sharp”) into otherwise Cantonese utterances that describe what can be seen in the book. The remaining four code-mixed utterances are: “See *kohko* ‘that’ cat”, “*Ikoh* ‘this’ kitty kissing”, “*Tai hui leongko* ‘see the two of them’ kiss kiss *haimou* ‘right’”, “*Nah ikoh* ‘here this’ kitty is running”. The Cantonese insertions are demonstratives (and deictic) to encourage the child to look at the pictures while the story is commented on in English. In turn 1, the English word “cat” is repeated eight times as the main topic of the book, followed by its synonym “kitty”, used in subsequent turns. PGM associates the actions depicted in the book to the grandfather’s action of eating in the same room and compares RYA’s eyes to the cat’s eyes. In turn 2 after the child turns the page, she explains the English word “peeking” in Cantonese. When asked in Cantonese how the cat peeks, RYA shows his understanding by opening his eyes wide. In turn 3, PGM provides translation equivalents for “strolling” and “rolling” to explain the actions of the cat. For “hissing”, she explains in Cantonese why a cat would hiss and uses the opportunity to warn against disturbing cats who may bite. She does not provide a Cantonese equivalent for “kissing” (a known word) but does so for “running”, supplementing it by running her fingers along the sofa they are sitting on. PGM unconsciously and inadvertently supports her grandson’s emergent bilingualism using the same types of cues and practices Faltis (2019) has outlined for classroom code-switching practices by teachers. Antón et al. (2016) also found that language mixing in the classroom during concept acquisition is not detrimental to learning outcomes.

PGM often switched to Cantonese while reading books in English whenever RYA offered comments in Cantonese, thus encouraging and engaging him to interact in whichever language he preferred. The child generally accommodated his grandmother’s language choice patterns proportionally. Interestingly, striking similarities in code-switching patterns as shown in Figure 1 and Table 2 were also found from corpus data of bilingual adults and developing bilingual children from the same community. Phillips & Deuchar (2022: 74) suggest that their corpus findings support a usage-based approach whereby children “are acquiring the [code-switching] patterns of speech produced in the input available to them” as depicted in this case study.

Suzanne Quay

4.2 Qualitative differences between the emergent bilingual child's English and Cantonese utterances

What is surprising from the quantitative analyses is that this child who hardly spoke any English at preschool did so with his grandmother from the earliest session, and in general, matched his grandmother's language choices. Despite similar code-switching patterns as his grandmother, qualitative analyses show that his English utterances were much shorter and more limited in morphosyntactic structure than his Cantonese utterances. Until almost age five, most of his English utterances were: one-word nouns referring to animals, food or objects as depicted in books; various short phrasal utterances like "no more paper", "all finished", or "up the ladder"; or full or partial repetitions of his grandmother's English utterances. He was often asked to repeat English words in the preschool setting, and PGM did the same explicitly at home, for example at age 3;4.9, in "Big hug *nei kong la*" ('you say (it)'). All such utterances and some set phrases like "I don't know", "I don't remember", and "Wait a minute" inflated his English total. Towards age five, he produced a few spontaneous short sentences like: "I see this", "I think it's a worm", "I found it", "I rolled six", "Six is the biggest number". His longest English utterances occurred when singing songs like "Old MacDonald had a farm" or when reciting the alphabet or counting aloud.

Despite accommodating his grandmother's use of English from the first session, RYA was dominant in Cantonese, which is not obvious from the quantitative analyses. Table 3 lists 20 examples of his typical Cantonese utterances during the preschool period to show that in contrast to his English utterances, his Cantonese utterances were well-formed sentences with finite verbs from Session 1. The last column provides literal translations into English with parentheses around English elements not in the original Cantonese to show congruence where possible (when needed, a better English gloss is provided after the equal symbol between parentheses, and square brackets surround additional contextual information not expressed in original examples). Numbered examples in Cantonese are preceded by C in Table 3, Mixed ones by M in Table 4, and English examples by E in Table 5.

A brief description of Cantonese grammar is necessary. Cantonese, like other Chinese dialects, is a pro-drop language with very little of the grammatical morphology prominent in English, such as inflectional markings of person, gender, case, number (except for pronouns), tense or mood (Matthews & Yip 2011). Indeed, Cantonese has few affixes compared to English; its grammatical relationships are mainly encoded by relative word position and strict adherence to word order. Like English, its basic word order is SVO (subject-verb-object) as in C4, C8–

7 Supporting bilingual development through code-switching

9, C17–18 in Table 3. But unlike English, the following deviations are possible: (7) SOV as in C12; (8) V + S order or subject-verb inversion; (9) right-dislocation where an element is put last or added separately from the rest of the sentence (as in M5, Table 4); and (10) topicalization (as in M6). Because Cantonese is a topic prominent language, it has a more flexible word order than English since various elements of a sentence can become the sentence topic by being placed first (Matthews & Yip 2011). Implicit subjects are identified by pragmatic-discourse criteria from the immediate speech context and discourse topic (Luke et al. 2003). This is evident from examples C2, C3, C5–7, C10, C14–16, and C20. While every sentence must have a subject in English, this is not required in Cantonese as indicated by the parentheses surrounding overt subjects in English translations of RYA's Cantonese (Table 3) and ENG-CHI compound/complex utterances (Table 4). The subject pronoun can be retrieved from the immediate speech context as for C6 which RYA said while looking at a turkey being roasted in the oven. In C15, he is at the table with PGM when he refers to how the alphabet letters he is writing extends to where his grandmother is sitting. The preposition (PREP) *hai* in C4 and C5 function like verbs and can occur in a sentence on its own without a verb as in those two examples. RYA's well-formed Cantonese utterances conveyed his intentions clearly as when he told PGM in C9 that he would change his pants after spilling water on himself or in C4 at 2;9.25 when he informed PGM that he would fetch a toy telephone for them to play with. C17 and C20 have the aspect marker (ASP) indicating past tense.

Table 3 does not show the production of all possible grammatical elements in Cantonese. For example, while PGM produced sentence final particles (SFP) like *la* in *nei kong la* 'you say (it)' or *lo* in lines 5–6 of Example (9), none appeared in RYA's Cantonese utterances. RYA does have some preposition (PREP in C4–5), particle (PRT in C6, C18) and aspect (ASP in C17, C20) markers. Note that Cantonese-speaking preschool children in Hong Kong also develop grammatical features at different rates according to age and individual factors (Wong 2023).

4.3 Discourse strategies in multi-word Mixed and English utterances

Example (8) is a typical dual-language conversation. At age 4;8.19, RYA is almost matching PGM here in number of turns (4RYA:5PGM) in contrast to 15 months earlier in Example (7) (1RYA:3PGM). PGM asks the first question in Cantonese but the next two Cantonese questions contain a conjunction and proper name (of a shopping mall) in English. RYA then responds with a Cantonese deictic expression, *kohko*, followed in English by the conjunction "and" with the mall's name. In lines 5 to 12, when the conversation is relatively simple and straightforward,

Suzanne Quay

Table 3: Examples of Ryan’s Cantonese utterances

Age (Sess. #)	Ex	Cantonese	Literal translations (ENG elements absent in CAN in parentheses)
2;9.25 (1)	C1	<i>Itit um nau nau nau nau.</i>	This (does) not bite bite bite bite.
	C2	<i>Yap umtuk hui.</i>	Inside can’t go (=It) can’t go in).
	C3	<i>Um pei chueyfan.</i>	(You are) not allowed to cook.
	C4	<i>Ngo lo tinwa hai kohtou.</i>	I bring the telephone PREP (from) there.
3;4.9 (4)	C5	<i>Chung yau ye hai yappin.</i>	Still more things PREP inside (it) (=There are) more things inside).
	C6	<i>Oi chuey dou hui hui um yuk tuk.</i>	(You) must cook PRT (until [turkey]) can’t move.
	C7	<i>Tit zo yap hui.</i>	(It) fell inside.
	C8	<i>Ngo um oi tuk.</i>	I (do) not want to read (it).
3;7.24 (8)	C9	<i>Ngo hui wun jek tiew foo.</i>	I (will) go (to) change (my) pants.
	C10	<i>Lau pei shuey.</i>	drip nose water (=I have a runny nose).
	C11	<i>Maamaa hui pintou?</i>	Grandma go where? (=where is Grandma going?)
	C12	<i>Ijek yan mou chin mai.</i>	This person (does) no (=not-have) money (to) buy/spend.
4;4.10 (12)	C13	<i>Hai meye ikoh?</i>	Is what this? (=What is this?)
	C14	<i>Hai kum yong.</i>	(It) is like this.
	C15	<i>Cheung tou maamaa kohtou.</i>	Long PREP (= [letters he’s writing] reaches) grandma over-there.
4;10.14 (16)	C16	<i>Kum hakmanman.</i>	(It’s) so dark.
	C17	<i>Ryan um kei tuk zo.</i>	Ryan (does) not remember ASP.
	C18	<i>Maamaa tai um dou.</i>	Grandma (could) not see PRT.
	C19	<i>Hai mai itou?</i>	Is-or-is-not (it) here?
	C20	<i>Choh zo.</i>	Wrong ASP (=It was wrong)].

7 Supporting bilingual development through code-switching

both PGM and RYA continue speaking in English. He answers his grandmother in line 9 with one rare full English sentence, “I like Metrotown too.” Line 11 (“Every sushi store”) is more typical of his short phrasal utterances in English, which his grandmother expands into a full sentence in line 12.

- (8) Ryan and his grandmother discuss which sushi places he likes at age 4;8.19
- 1 *PGM: *Kumyat sushi housek mah* ‘was the sushi yummy today’?
- 2 *PGM: Uh?
- 3 *PGM: *Nei chungyi kumyat sek kohkan* or *chungyi hui Metrotown sek kohkan* ‘do you like the place we ate at today or do you like the one at Metrotown’?
- 4 *RYA: *Kohko* ‘that one’ and Metrotown.
- 5 *PGM: Both?
- 6 *PGM: You also like today’s one?
- 7 *PGM: Davie one [street name].
- 8 *PGM: You don’t like Metrotown?
- 9 *RYA: I like Metrotown too.
- 10 *PGM: Which one is better?
- 11 *RYA: Every sushi store.
- 12 *PGM: Every sushi store is good?
- 13 *RYA: Metrotown upstairs *yau keh* ‘have that’ store.
- 14 *RYA: Down *hai* ‘is’ ... upstairs *yau keh* ‘have that’ sushi place.
- 15 *RYA: And down there *yau* ‘have’...and and...up there *yau* sushi place.
- 16 *RYA: And down there *yau* sushi place.
- 17 *PGM: *uh koh* ‘that’ foodcourt *lo*.
- 18 *RYA: *yau* three, three sushi place.

He code-switched intrasententially when he needed to convey more complicated information (lines 13–16 and 18) about the location of three different Japanese restaurants in the mall. Sharing the same background knowledge of the layout of the mall, PGM understood him even though his elaborations consist of Cantonese verbal elements with seemingly truncated noun and prepositional phrases in English.

Combining English and Cantonese in multiword utterances is the child’s typical discourse strategy to extend the messages he can convey. Notably, his multiword ENG-CHI utterances were the longest, comprising 30% of all his intrasentential code-switched utterances ($n = 211$, Table 2). Although his Cantonese-only utterances were simple sentences, some of his multiword *mixed* utterances were

Suzanne Quay

complex sentences (with coordinate or subordinate clauses) due to congruent lexicalization, an issue rarely discussed (as mentioned earlier) in the literature on children's code-switching (Treffers-Daller 2022). Muysken (1997, 2000) in describing this type of code-switching for adults, explained that it occurs when two languages share grammatical structures that can be filled lexically with elements from either language. Example (9) shows that RYA produced an 11-word ENG-CHI utterance equivalent in English to a complex sentence with a subordinating conjunction in line 4:

- (9) Reading a book at age 3;4.9
 1 *PGM: E for?
 2 *RYA: Egg.
 3 *PGM: Oh, and it is cracking.
 4 *RYA: baby *nganngan tuk chut lei tai teiha yau mou mama* birdie 'the baby [bird] is looking out to see if its mother birdie is on the ground'.
 5 *PGM: *hai lo* 'yes'.
 6 *PGM: *tai hui mama hai tou mou lo* 'it is checking if its mother is around'.

Predominantly Cantonese mixed utterances as in line 4 above were much longer than predominantly English mixed utterances. While RYA produced more English words in Example (8) at 4;8.19 (lines 13–16, 18), the insertion of verbs in Cantonese indicate a reliance on Cantonese morphosyntactic structure. While *yau* 'have' was used the most often, other Cantonese verbs often inserted into English utterances are *hai* 'is/copula', *hui* 'go', *kiew* 'is called' and *pei* 'give'. Cantonese verb insertions helped the child to produce relatively long English utterances early on:

- (10) Looking at alphabet flashcards at age 3;4.9
 1 *RYA: And *yau* 'have' one two three flower in there.
 2 *PGM: Three flowers. Wow. That's good.

Pointing out three flowers on the flashcard, he inserts English words into a Cantonese framework where 'there' occurs at the end as in *yau yat yi sam-go faa hai godou* and not at the beginning as it would in an English utterance. Although PGM models the plural inflection of "flower" in line 2, RYA's lack of plural marking is also a sign of congruent lexicalization as his more established Cantonese system does not require such inflections nor subject pronouns.

Table 4 lists 11 examples from 54 recordings (including the 16 sessions used earlier) of ENG-CHI complex and compound utterances that provide evidence

7 Supporting bilingual development through code-switching

that: (7) English elements are superimposed on Cantonese structures (i.e., congruent lexicalization), and (8) bilingual mixed utterances are much longer than unilingual English or Cantonese utterances. Although limited, the hybrid examples in Table 4 indicate congruent lexicalization and convergence occurring. For example, in M1, the English insertions fit the dependent clause preceding the independent one conveyed by ‘when this one is finished, I will use blue’ (referring to the ink in coloring pens). Both clauses, however, are derived from Cantonese verbs – *mou sai* ‘to be finished’ does not require the subordinating conjunction (when) and *jung* ‘use’ does not require a subject pronoun. Thus, “this one *mou saai zau jung* blue” is an example of the way Clyne (2003) defined convergence as a hybrid form that exhibits elements from two linguistic systems. The final column contains literal translations of the ENG-CHI utterances with parentheses around elements absent in Cantonese. If all the English insertions are translated directly into Cantonese from M1–M7 and M9–M11, he would have well-formed Cantonese utterances. For example, in relating how he won a game of Snakes & Ladders, RYA’s English insertion of “number and then win” in M3 is a direct translation from Cantonese of *souzi tungmaai jengzo* that follows after *hondo* ‘a lot of’ without the plural inflection on “number”, without the first-person subject pronoun (I) or past tense inflection (albeit the irregular “won”), all of which are not required in Cantonese and reflected in his English lexical insertions. M8 is the only example that does not share complete word-order congruence in the two languages as the placement of “more” follows the English order after ‘a little’ rather than before in Cantonese as in *ngo zing do siusiu*. The subordinating conjunction *mai* ‘so that’ is followed by the English verb “make” with a Cantonese direct object in the same way as a literal Cantonese translation [*choudou*] *hondo chin*. Gawlitzek-Maiwald & Tracy (2005: 277) proposed that “convergence comes about when children (assisted by universal grammar) reconstruct derivational relationships linking hitherto independent systems” and that this gives us insights into children’s early awareness of cross-linguistic equivalence and contrasts. The child used this transitional strategy for more than a year from M1 to M11 because his grandmother never failed to understand him. For example, M7 was produced when he asked his grandmother about a book with text he had seen previously but could not find that day. She told him the book had already been returned to the library. Her response conveyed to him the acceptability of his code-mixing pattern which did not impede their meaningful communication.

Although PGM modelled mixed utterances, she produced a smaller proportion at 11% versus RYA at 19% (Figure 1) and decreased such utterances greatly through adherence to intersentential code-switching by the last recording while RYA’s intrasentential code-switching increased. Only in the last session does he produce

Suzanne Quay

three English-based mixed utterances with one-word Cantonese insertions as shown below (excluding the kinship term in b), indicating perhaps the start of a shift in his Cantonese dominance (although M11 in Table 4 still shows Cantonese verbs supporting predominantly English utterances). He produces:

- (11) ENGchi utterances at 4;10.14
- a. I think Ryan roll *zo*.
 - b. Ryan move *zo maamaa* down.
 - c. How Ryan try the red one *yausi*?

The past tense of English verbs is indicated by the Cantonese aspect marker *zo* or ‘already’ rather than with an ‘-ed’ inflection in a and b. He also used the split order of the verb-particle (“move down”), preferred in English but not in Cantonese. Yip & Matthews (2016: 12) considered this an innovation when their Cantonese-dominant children in Hong Kong produced the split order in Cantonese-based mixed utterances not found in the adult Cantonese input. Likewise, a and b are an innovative use of congruent lexicalization for RYA in Canada with English as the base for Cantonese insertions since PGM never used *zo* with an English verb but did produce it as expected with Cantonese verbs in mixed utterances like “The boat *lan zo* ‘is broken’” (i.e., qualitatively different from RYA’s mixing). In c, the utterance is mostly English except for the Cantonese adverb *yausi* (‘again’), indicating an emerging English proficiency.

RYA’s well-formed Cantonese utterances (Table 3) provide evidence of his language dominance which influenced not only English insertions in ENG-CHI complex/compound utterances (Table 4) but also the production of unilingual English utterances. Previous studies have focused on code-mixed utterances without examining utterances produced in a single code. If we accept that the child is using his stronger language to support his weaker one through congruent lexicalization, we can examine convergence or hybridity in his spontaneously produced English utterances. As mentioned earlier, very few of his English utterances contained subjects and predicates. Most of his two-or-more-word ENG utterances appear to be ungrammatical in English when analyzed according to English grammatical rules but not when analyzed according to Cantonese grammatical rules. Table 5 shows 20 such *ungrammatical* English examples translated into Cantonese. The literal Cantonese translations with pro-drop/subject omission in E1–5, E9–11, E19–20 show that the English utterances were based on the morphosyntactic structure of Cantonese (see Wong 2023 for more on Cantonese grammar). Unlike the Hong Kong bilingual child who was also Cantonese-dominant in Yip

7 Supporting bilingual development through code-switching

Table 4: The child's ENG-CHI utterances beyond simple sentences

Age	Ex	ENG-CHI (or Mix beyond one-word insertions)	Literal translations (ENG elements absent in CAN in parentheses)
3;5.27	M1	This one <i>mou saai zau jung</i> blue.	(When) this one (is) <i>finished</i> (I will) <i>use</i> blue.
3;9.28	M2	How <i>yeye chou</i> chair <i>tai tv</i> and then Ryan <i>dyeu bei yeye hai</i> chair.	How(-about) <i>grandpa sit(s) on</i> (the) chair (to) <i>watch</i> TV and then Ryan <i>throws</i> [toy] <i>to grandpa on</i> (the) chair.
4;1.23	M3	<i>Gumjat</i> Ryan <i>houdo</i> number and then win.	<i>Today</i> Ryan (had) (a) <i>lot of</i> number(s) and then (I won).
4;2.13	M4	<i>Maamaa hai mai</i> Ryan <i>jau</i> nine and then go up the ladder?	<i>Grandma is-or-is-not</i> (that when) Ryan <i>gets</i> nine and then go(es) up the ladder?
4;3.17	M5	<i>Heoi go godou sin</i> turn remember?	<i>After we reach there</i> , (we) turn, remember?
4;4.17	M6	Ryan two year old <i>ngo lei</i> and then <i>cheng lan hai mai</i> ?	(When) Ryan (was) two year(s) old <i>I came</i> and then <i>broke</i> (it) <i>didn't</i> (I)?
4;4.10	M7	<i>Maamaa</i> how come last time <i>jau go book hai godou jau word se hai godou</i> .	<i>Grandma</i> , how come last time <i>there was</i> (a) book (and there) <i>were</i> word(s) <i>written there</i> .
4;7.30	M8	<i>Ngo zing siusiu</i> more <i>mai</i> make <i>houdo chin</i> .	<i>I make a little more so that</i> (I can) make <i>lots of money</i> . [inverted: little more= <i>do siusiu</i>]
4;7.30	M9	How [Ryan] push this up and then <i>yeye faaidit se</i> .	How(-about) Ryan push this up and then grandpa hurry (to) shoot.
4;8.19	M10	How <i>maamaa sai zo wun maamaa tung ngo waan</i> snake and ladder?	How(-about) [when] <i>grandma finish(es) washing</i> (the) <i>dishes</i> , <i>grandma play(s)</i> snake(s) and ladder(s) <i>with me</i> .
4;10.14	M11	<i>Hai</i> fifty-seven plus twelve <i>mai</i> sixty nine?	<i>Is</i> (it) fifty-seven plus twelve (that) <i>make(s)</i> sixty nine?

Suzanne Quay

& Matthews (2000), RYA always produced *wh*-words like *what* (E6), *where* (E12) and *how* (E13, 16, 19) consistently in initial position as required in English but not in Cantonese, a *wh*-in-situ language where *wh* words may serve as a subject or an object (Wong 2023). This interrogative construction may be more salient in English exposure in Canada than in Hong Kong and was the only one where RYA did not transfer Cantonese structure into his English utterances.

Although Cantonese and English are not closely related, both share a basic SVO structure that facilitated the transfer of the morphosyntactic frame of RYA's stronger language, Cantonese, into his English utterances, resulting in examples of congruent lexicalization and convergence in Tables 4 and 5. This was not apparent until English utterances were translated back into Cantonese. Thus, the surface English elements mask the underlying Cantonese morphosyntactic rules used in their construction, exemplifying clearly how separate languages can no longer be identified and is only "discernible after the fact; that is, convergence is an outcome (of borrowing or substratum influence, for instance) and not an observable process" (Bullock & Toribio Almeida 2004: 91). This use of hybrid morphosyntactic forms as translanguaging practice gives the child time to develop his English ability and acquire target English structures without compromising meaningful communication with his bilingual grandmother. His English grammar does develop over time with the production of articles and plural inflections in later sessions. For example, "Stir chocolate" (E3) lacks a definite article unnecessary in Cantonese that is produced later at 3;7.24 in E10–12 and E17; "Two flower" (E5) shows no plural inflection but a plural marking appears, albeit erroneously, in "This one twos" (E14). As mentioned earlier, prepositions in Cantonese can function like verbs and can occur in sentences on their own without verbs. This flexible usage of prepositions is exemplified with the preposition 'down' in E9–10, and E17 (also seen earlier in Example (8), lines 14–16). Although their results were based on *visible* code-switching, Yow et al. (2018: 1085; their emphasis) conclude in their study of preschool children in Singapore that "the ACT of code-switching by children may have provided them a way to engage both their languages more frequently, particularly the weaker language." Code-switching also allowed the Singaporean children to improve their developing bilingual proficiency as it does for RYA. Although PGM's utterances and RYA's utterances were qualitatively different, no breakdown in communication ever occurred, thus emphasizing that dual language use in the home was a *fluid* bilingual practice that gave the emergent bilingual child time to develop his weaker language.

7 Supporting bilingual development through code-switching

Table 5: Examples of Ryan’s English utterances

Age (Sess. #)	Ex	English	[= literal CAN translation]: English meaning
2;9.25 (7)	E1	No this one.	[=Umhai ikoh]: ‘(it’s) not this one’.
	E2	Wash like this.	[=Sai kumyeong]: ‘(I) wash (it) like this’.
	E3	Stir chocolate.	[=Gao zyugulik]: ‘(I) stir (the) chocolate’.
	E4	Water right.	[=Sui haimai]: ‘(it’s) water, isn’t it?’
3;4.9 (4)	E5	Two flower.	[=Loenggo faa]: ‘(there’s) two flower(s)’.
	E6	What behind?	[=Houmin meye]: ‘what (is) behind?’
	E7	There two.	[=Godou loenggo]: ‘there (is) two (of them)’.
	E8	This one ghost.	[=Jigo gwai]: ‘this-one (is a) ghost’.
3;7.24 (8)	E9	How down hill?	[=Dim loksaan]: ‘How (do I go) down (the) hill?’
	E10	How down the tail?	[=Dim lokheoi tiu mei]: ‘How (do I go) down (the) tail?’
	E11	Go up the ladder.	[=Soeng lautai]: ‘(I) go up (the) ladder’.
	E12	Where the paddle?	[=Syun joeng bindou]: ‘where (is the) paddle?’
4;4.10 (12)	E13	How about this one?	[=Jigo dim]: meaning ‘how (is) this one?’
	E14	This one twos.	[=Jigo leong go]: ‘this one (has) two (of them)’.
	E15	This one back too.	[=Jigo dou faanlei]: ‘this one (is) turned around’.
	E16	How this one?	[=Jigo dim]: ‘how (is) this one?’
4;10.14 (16)	E17	Down the snake.	[=Lok gogo se]: ‘(I go) down that snake’
	E18	More there.	[=Godou zungyau]: ‘(there’s) more there’.
	E19	How now?	[=jigaa dim]: ‘how (is it) now?’
	E20	How change color?	[=dim wun ngaansik]: ‘how (do I) change color?’

Suzanne Quay

5 Conclusion

Examining the child's code-switching practices revealed his abilities rather than the deficiencies often highlighted by the monolingual-normed understanding of proficiency subscribed by schools and society. The grandmother socialized the child into using his whole linguistic repertoire to express himself fully by modelling dual-language usage that made it acceptable for the child to produce mixed utterances. As a result of such bilingual interactions, the child was able to accommodate to his grandmother's use of English. Unlike his grandmother's data, he used congruent lexicalization (Muysken 1997, 2000) and convergence (Bullock & Toribio Almeida 2004, Clyne 2003, Toribio 2004) as part of his code-switching practices: (7) to convey information in longer mixed complex/compound sentences beyond the simple sentences of unilingual utterances, and (8) to form complete utterances in his weaker language by using the morphosyntactic rules of his stronger language. Both types of code-switching strategies where the child arranged English-only and English-mixed utterances according to Cantonese morphosyntactic rules reflect hybridity. This study reveals the need for researchers to examine not just intrasentential mixed utterances in future studies but also *unilingual* utterances in the weaker language where the code-mix through convergence may be invisible. The child in this study capitalizes on the structural congruence between his two languages. He uses congruent lexicalization and convergence as part of his strategy to accommodate his grandmother's use of English as well as insertion and language alternation that more clearly demarcate his two languages. RYA's use of congruent lexicalization/convergence in his code-switching practice provides evidence not only that his Cantonese system is stronger than his English system but that he can use bilingual bootstrapping (Gawlitzeck-Maiwald & Tracy 1996) and cross-linguistic lexical and structural equivalence to increase the production of his weaker language. This mixing/hybrid pattern reveals also what the child may perceive as more versus less advanced options as his weaker language catches up with the stronger one in development. What is unclear is whether longitudinal empirical data can indicate how he distinguishes his coexisting structural formats from coexisting grammars and where he draws the limit between these options. That is, when does his strategy of convergence yield to two distinct systems? This is not apparent from the data set used for this study. In personal communication with the child's mother, it appears that RYA was assessed by his school around age 6;6 as on a par with monolingual English peers, which suggests that his grandmother's code-switching practices helped him to transition faster from the home to the school language.

7 *Supporting bilingual development through code-switching*

Case studies, like this one, are valuable (despite their obvious limitation) because they can refute generalizations, show what is possible, and give rise to hypotheses (as explained in Deuchar & Quay 2001). This study refutes the generalization that emergent bilinguals are producing *broken* English when they enter school. The hybrid nature of RYA's English utterances (in his use of congruent lexicalization and convergence) indicate child agency because despite limited English ability, he used his available linguistic resources to match his grandmother's language choice. Indeed, the child produced a similar proportion of English-only utterances as his grandmother from the outset and could form much longer complex/compound sentences when he could draw on his two languages. From a methodological viewpoint, this case study reveals that code-switching research should not just focus on dual-language utterances but also on single-language utterances that appear ungrammatical in a weaker language since convergence as a substratum influence may not be observable at the surface-level of the utterance (Bullock & Toribio Almeida 2004). That is, English utterances can be analyzed not only with English grammar rules but also with the grammar of children's stronger language to consider whether congruency allows that grammar to be used as a base for the weaker language.

This study has implications not only for raising bilingual children at home using code-switching input but also for the treatment of emergent bilinguals in the school setting. Primary school teachers need to appreciate that children can draw upon linguistic resources from their HLs to develop a weaker school language and not view emergent bilinguals as linguistically deficient. Emergent bilinguals or English language learners (ELLs) are expected to make up 25% of school-aged children in the US by 2025 (National Education Association 2020) with many natively-born rather than newcomers (Zong & Batalova 2015). More research is needed to address the dilemma mentioned earlier of when and how to introduce the societal language without compromising HL maintenance, which Wong Fillmore (1991: 345) states is related to "timing, not English" adding that the "children have to learn English, but they should not be required to do so until their native languages are stable enough to handle the inevitable encounter with English and all it means." Three decades after that exhortation, we still do not know when a HL can be considered stable enough for it to survive when these children start English medium education. Clearly, legitimizing HL identity and changing attitudes toward an acceptance of hybrid language practices for ELLs would encourage and validate the maintenance of HLs in English-speaking countries.

Suzanne Quay

Acknowledgment

I am enormously grateful to the child's parents and extended family for being a part of this case study, but especially to PGM and RYA, without whom this study would not have been possible. Thanks are also due to Janice Nakamura and Sheung Yu Cheung for their assistance with the transcriptions, coding, and translations. I am also indebted to the anonymous reviewer whose extensive comments helped greatly to reshape the focus of this chapter.

References

- Alba, Richard, John Logan, Amy Lutz & Brian Stults. 2002. Only English by the third generation? Loss and preservation of the mother tongue among the grandchildren of contemporary immigrants. *Demography* 39(3). 467–484.
- Antón, Eneko, Guillaume Thierry, Alexander Gborov, Jon Anasagasti & Jon Andoni Duñabeitia. 2016. Testing bilingual educational methods: A plea to end the language-mixing taboo. *Language Learning* 66(S2). 29–50. DOI: [10.1111/lang.12173](https://doi.org/10.1111/lang.12173).
- Balam, Osmer. 2021. Beyond differences and similarities in codeswitching and translanguaging research. *Belgian Journal of Linguistics* 35(1). 76–103. DOI: [10.1075/bjl.00065.bal](https://doi.org/10.1075/bjl.00065.bal).
- Braun, Andreas. 2012. Language maintenance in trilingual families – A focus on grandparents. *International Journal of Multilingualism* 9(4). 423–436. DOI: [10.0718.2012.714384](https://doi.org/10.0718.2012.714384).
- Bullock, Barbara E. 2009. Phonetic reflexes of code-switching. In Almeida Jacqueline Toribio & Barbara E. Bullock (eds.), *The Cambridge handbook of linguistic code-switching* (Cambridge Handbooks in Language and Linguistics), 163–181. Cambridge: Cambridge University Press.
- Bullock, Barbara E. & Jacqueline Toribio Almeida. 2004. Introduction: Convergence as an emergent property in bilingual speech. *Bilingualism: Language and Cognition* 7(2). 91–93. DOI: [10.1017/S4001506](https://doi.org/10.1017/S4001506).
- Cantone, Katja F. 2022. Language exposure in early bilingual and trilingual acquisition. *International Journal of Multilingualism* 19(3). 402–417. DOI: [10.0718.2019.1703995](https://doi.org/10.0718.2019.1703995).
- Cenoz, Jasone & Durk Gorter. 2022. Pedagogical translanguaging and its application to language classes. *RELC Journal* 53(2). 342–354. DOI: [10.1177/003368822210827](https://doi.org/10.1177/003368822210827).
- Chevalier, Sarah. 2013. Caregiver responses to the language mixing of a young trilingual. *Multilingua* 32(1). 1–32. DOI: [10.1515/multi-2013-0001](https://doi.org/10.1515/multi-2013-0001).

7 Supporting bilingual development through code-switching

- Chondrogianni, Vasiliki & Evangelia Daskalaki. 2023. Heritage language use in the country of residence matters for language maintenance, but short visits to the homeland can boost heritage language outcomes. *Frontiers in Language Sciences* 2. 1230408. DOI: [10.3389/flang.408.1230408](https://doi.org/10.3389/flang.408.1230408).
- Clyne, Michael. 2003. *Dynamics of language contact: English and immigrant languages*. 1st edn. Cambridge: Cambridge University Press. DOI: [10.1017/CBO9780511606526](https://doi.org/10.1017/CBO9780511606526).
- Comeau, Liane, Fred Genesee & Lindsay Lapaquette. 2003. The Modeling Hypothesis and child bilingual codemixing. *International Journal of Bilingualism* 7(2). 113–126. DOI: [10.1177/13670069030070020101](https://doi.org/10.1177/13670069030070020101).
- Comeau, Liane, Fred Genesee & Morton Mendelson. 2007. Bilingual children's repairs of breakdowns in communication. *Journal of Child Language* 34(1). 159–174.
- Curdt-Christiansen, Xiao Lan. 2013. Implicit learning and imperceptible influence: Syncretic literacy of multilingual Chinese children. *Journal of Early Childhood Literacy* 13(3). 348–370. DOI: [10.7984/12455819](https://doi.org/10.7984/12455819).
- Curdt-Christiansen, Xiao Lan & Baoqi Sun. 2022. Establishing and maintaining a multilingual family language policy. In Anat Stavans & Ulrike Jessner (eds.), *The Cambridge handbook of childhood multilingualism*, 257–277. Cambridge: Cambridge University Press. DOI: [10.1017/69771](https://doi.org/10.1017/69771).
- De Houwer, Annick. 2020. Harmonious bilingualism: Well-being for families in bilingual settings. In Andrea C. Schalley & Susana A. Eisenchlas (eds.), *Handbook of home language maintenance and development: Social and affective factors*, 63–83. Berlin: De Gruyter Mouton. DOI: [10.1515/9175-005](https://doi.org/10.1515/9175-005).
- Deuchar, Margaret. 2013. Code Switching. In Carol A. Chapelle (ed.), *The encyclopedia of applied linguistics*. Malden: Blackwell. DOI: [10.1002/9781405198431.wbeal0142](https://doi.org/10.1002/9781405198431.wbeal0142).
- Deuchar, Margaret. 2022. Multilingualism in early childhood: The role of the input. In Anat Stavans & Ulrike Jessner (eds.), *The Cambridge handbook of childhood multilingualism*, 58–81. Cambridge: Cambridge University Press. DOI: [10.1017/69771.005](https://doi.org/10.1017/69771.005).
- Deuchar, Margaret & Suzanne Quay. 1999. Language choice in the earliest utterances: A case study with methodological implications. *Journal of Child Language* 26(2). 461–475. DOI: [10.1017/S0305000999003852](https://doi.org/10.1017/S0305000999003852).
- Deuchar, Margaret & Suzanne Quay. 2001. What use is a case study of bilingual acquisition? In Margareta Almgren, Andoni Barreña, María-José Ezeizabarrena, Itziar Idiazabal & Brian MacWhinney (eds.), *Research on child language acquisition: Proceedings of the 8th conference of the International Association for the Study of Child Language*, 128–136. Somerville: Cascadilla Press.

Suzanne Quay

- Eisenchlas, Susana A. & Andrea C. Schalley. 2020. Making sense of “home language” and related concepts. In Andrea C. Schalley & Susana A. Eisenchlas (eds.), *Handbook of home language maintenance and development: Social and affective factors*, 17–37. Berlin: De Gruyter Mouton. DOI: [10.1515/9175-002](https://doi.org/10.1515/9175-002).
- Faltis, Christian J. 2019. Pedagogical codeswitching and translanguaging in bilingual schooling contexts: Critical practices for bilingual teacher education. In Jeff MacSwan & Christian J. Faltis (eds.), *Codeswitching in the classroom: Critical perspectives on teaching, learning, policy, and ideology*, 39–62. New York: Routledge.
- García, Ofelia, Jo Anne Kleifgen & Lorraine Falchi. 2008. *From English language learners to emergent bilinguals*. New York: Campaign for Educational Equity. <https://eric.ed.gov/?id=ED524002>.
- García, Ofelia. 2009. *Bilingual education in the 21st century: A global perspective*. Chichester: Wiley-Blackwell.
- Gawlitzeck-Maiwald, Ira & Rosemary Tracy. 1996. Bilingual bootstrapping. *Linguistics* 34(5). 901–926. DOI: [10.1515/ling.1996.34.5.901](https://doi.org/10.1515/ling.1996.34.5.901).
- Gawlitzeck-Maiwald, Ira & Rosemary Tracy. 2005. The multilingual potential in emerging grammars. *International Journal of Bilingualism* 9(2). 277–297.
- Genesee, Fred. 2022. Evidence for differentiated languages from studies of bilingual first language acquisition. In Jeff McSwan (ed.), *Multilingual perspectives on translanguaging*, 211–229. Bristol: Multilingual Matters.
- Gort, Mileidis. 2019. Young emergent bilinguals’ literate and languaging practices in story retelling. In Jeff MacSwan & Christian J. Faltis (eds.), *Codeswitching in the classroom: Critical perspectives on teaching, learning, policy, and ideology*, 162–183. New York: Routledge.
- Gregory, Eve, Tahera Arju, John Jessel, Charmian Kenner & Mahera Ruby. 2007. Snow white in different guises: Interlingual and intercultural exchanges between grandparents and young children at home in East London. *Journal of Early Childhood Literacy* 7(1). 5–25. DOI: [10.798407074831](https://doi.org/10.798407074831).
- Hoff, Erika & Cynthia Core. 2013. Input and language development in bilingually developing children. *Seminars in Speech and Language* 34(4). 215–226. DOI: [10.1055/s-448](https://doi.org/10.1055/s-448).
- Hulk, Aafke & Natascha Müller. 2000. Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition* 3(3). 227–244. DOI: [10.1017/S1366728900000353](https://doi.org/10.1017/S1366728900000353).
- Ishizawa, Hiromi. 2004. Minority language use among grandchildren in multi-generational households. *Sociological Perspectives* 47(4). 465–483. DOI: [10.1525/sop.2004.47.4.465](https://doi.org/10.1525/sop.2004.47.4.465).

7 Supporting bilingual development through code-switching

- Kenner, Charmian, Mahera Ruby, Eve Gregory, John Jessel & Tahera Arju. 2007. Intergenerational learning between children and grandparents in East London. *Journal of Early Childhood Research* 5(2). 219–243. DOI: [10.718X07080471](https://doi.org/10.718X07080471).
- King, Kendall, Lyn Fogle & Aubrey Logan-Terry. 2008. Family language policy. *Language and Linguistics Compass* 2(5). 907–922. DOI: [10.1111/j.1749-818X.2008.00076.x](https://doi.org/10.1111/j.1749-818X.2008.00076.x).
- Lanza, Elizabeth. 1997. *Language mixing in infant bilingualism: A sociolinguistic perspective*. Oxford: Clarendon.
- Lanza, Elizabeth & Kristin Vold Lexander. 2019. Family language practices in multilingual transcultural families. In Simona Montanari & Suzanne Quay (eds.), *Multidisciplinary perspectives on multilingualism: The fundamentals*, 229–252. Berlin: De Gruyter Mouton. DOI: [10.1515/9984-011](https://doi.org/10.1515/9984-011).
- Lanza, Elizabeth & Rafael Lomeu Gomes. 2020. Family language policy: Foundations, theoretical perspectives and critical approaches. In Andrea C. Schalley & Susana A. Eisenclas (eds.), *Handbook of home language maintenance and development: Social and affective factors*, 153–173. Berlin: De Gruyter Mouton. DOI: [10.1515/9781501510175-008](https://doi.org/10.1515/9781501510175-008).
- Liang, Feng & Dong-Shin Shin. 2021. Heritage language maintenance of Chinese immigrant families: Perceptions, practices, and challenges. *Bilingual Research Journal* 44(1). 23–38. DOI: [10.5882/2021.1922539](https://doi.org/10.5882/2021.1922539).
- Little, Sabine. 2020a. Social media and the use of technology in home language maintenance. In Andrea C. Schalley & Susana A. Eisenclas (eds.), *Handbook of home language maintenance and development: Social and affective factors*, 257–273. Berlin.
- Little, Sabine. 2020b. Whose heritage? What inheritance?: Conceptualising family language identities. *International Journal of Bilingual Education and Bilingualism* 23(2). 198–212. DOI: [10.0050.2017.1348463](https://doi.org/10.0050.2017.1348463). (1080).
- López, Luis. 2020. *Bilingual grammar: Toward an integrated model*. Cambridge: Cambridge University Press.
- Luke, Kang Kwong, Bodomo Adams B. & Owen T. Nancarrow. 2003. On the subject condition in Cantonese. *Journal of Chinese Linguistics* 19. 32–57. <https://www.jstor.org/stable/23825489>.
- MacSwan, Jeff. 2016. Code-switching in adulthood. In Elena Nicoladis & Simona Montanari (eds.), *Bilingualism across the lifespan: Factors moderating language proficiency*, 183–200. Washington, D.C. & Berlin: American Psychological Association (APA) & De Gruyter Mouton.
- MacSwan, Jeff. 2019. Sociolinguistic and linguistic foundations of codeswitching research. In Jeff MacSwan & Christian J. Faltis (eds.), *Codeswitching in the*

Suzanne Quay

- classroom: Critical perspectives on teaching, learning, policy, and ideology*, 3–38. New York: Routledge.
- MacSwan, Jeff. 2022. Codeswitching, translanguaging and bilingual grammar. In Jeff MacSwan (ed.), *Multilingual perspectives on translanguaging*, 109–152. Bristol: Multilingual Matters.
- MacSwan, Jeff, Natalia Guzman, Kara T. McAlister & Margaret Marcus. 2019. Effects of home codeswitching practices on bilingual language acquisition. In Jeff MacSwan & Christian J. Faltis (eds.), *Codeswitching in the classroom: Critical perspectives on teaching, learning, policy, and ideology*, 148–161. New York: Routledge.
- Matthews, Stephen & Virginia Yip. 2011. *Cantonese: A comprehensive grammar*. 2nd edn. New York: Routledge.
- Meddegama, Indu Vibha. 2020. Cultural values and practices: The pillars of heritage language maintenance endeavours within an immigrant multilingual malayali community in the UK. *International Journal of Bilingual Education and Bilingualism* 23(6). 643–656. DOI: [10.0050.2019.1701978](https://doi.org/10.10050.2019.1701978).
- Meisel, Jürgen M. 1994. Code-switching in young bilingual children: The acquisition of grammatical constraints. *Studies in Second Language Acquisition* 16(4). 413–439. DOI: [10.1017/S0272263100013449](https://doi.org/10.1017/S0272263100013449).
- Melo-Pfeifer, Silvia. 2015. The role of the family in heritage language use and learning: Impact on heritage language policies. *International Journal of Bilingual Education and Bilingualism* 18(1). 26–44. DOI: [10.1080/13670050.2013.868400](https://doi.org/10.1080/13670050.2013.868400).
- Montanari, Simona. 2009. Pragmatic differentiation in early trilingual development. *Journal of Child Language* 36(3). 597–627. DOI: [10.1017/S0305000908009112](https://doi.org/10.1017/S0305000908009112).
- Montrul, Silvina. 2022. The development of the heritage language in childhood bi-/multilingualism. In Anat Stavans & Ulrike Jessner (eds.), *The Cambridge handbook of childhood multilingualism*, 537–554. Cambridge: Cambridge University Press. DOI: [10.1017/69771.029](https://doi.org/10.1017/69771.029).
- Mu, Guanglun Michael. 2014. Learning Chinese as a heritage language in Australia and beyond: The role of capital. *Language and Education* 28(5). 477–492. DOI: [10.1080/09500782.2014.908905](https://doi.org/10.1080/09500782.2014.908905).
- Muysken, Pieter. 1997. Code-switching processes: Alternation, insertion, congruent lexicalization. In Martin Pütz (ed.), *Language choices: Conditions, constraints, and consequences*, 361–380. Amsterdam: John Benjamins.
- Muysken, Pieter. 2000. *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.

7 Supporting bilingual development through code-switching

- Myers-Scotton, Carol. 1993. *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon.
- Nakamura, Janice. 2020. Language regrets: Mixed-ethnic children's lost opportunity for minority language acquisition in Japan. *Multilingua* 39(2). 213–237. DOI: [10.1515/multi-2019-0040](https://doi.org/10.1515/multi-2019-0040).
- Nakamura, Janice & Suzanne Quay. 2012. The impact of caregivers' interrogative styles in English and Japanese on early bilingual development. *International Journal of Bilingual Education and Bilingualism* 15(4). 417–434. DOI: [10.1080/13671433.2012.665827](https://doi.org/10.1080/13671433.2012.665827).
- National Education Association. 2020. *English language learners*. <https://www.nea.org/resource-library/english-language-learners>.
- Nicoladis, Elena. 2013. Code mixing. In Carol A. Chapelle (ed.), *The encyclopedia of applied linguistics*, 1–4. Malden: Blackwell. DOI: [10.1002/97831.wbeal0141](https://doi.org/10.1002/97831.wbeal0141).
- Palviainen, Åsa. 2020. Future prospects and visions for family language policy research. In Andrea C. Schalley & Susana A. Eisenclas (eds.), *Handbook of home language maintenance and development: Social and affective factors* (Handbooks of Applied Linguistics 18), 236–254. Berlin: De Gruyter Mouton. DOI: [10.1515/9781501510175-012](https://doi.org/10.1515/9781501510175-012).
- Park, Seong Man & Mela Sarkar. 2007. Parents' attitudes toward heritage language maintenance for their children and their efforts to help their children maintain the heritage language: A case study of Korean-Canadian immigrants. *Language, Culture and Curriculum* 20(3). 223–235. DOI: [10.2167/lcc337.0](https://doi.org/10.2167/lcc337.0).
- Phillips, Shannon & Margaret Deuchar. 2022. The role of the input in the acquisition of code-switching. In Unn Røyneland & Robert Blackwood (eds.), *Multilingualism across the lifespan* (Routledge Critical Studies in Multilingualism), 56–79. New York: Routledge. DOI: [10.4324/9781003125815-5](https://doi.org/10.4324/9781003125815-5).
- Purkarthofer, Judith. 2020. Intergenerational challenges: Of handing down languages, passing on practices, and bringing multilingual speakers into being. In Andrea C. Schalley & Susana A. Eisenclas (eds.), *Handbook of home language maintenance and development: Social and affective factors* (Handbooks of Applied Linguistics 18), 130–150. Berlin: De Gruyter Mouton. DOI: [10.1515/9781501510175-007](https://doi.org/10.1515/9781501510175-007).
- Quay, Suzanne. 2012. Discourse practices of trilingual mothers: Effects of minority home language development in Japan. *International Journal of Bilingual Education and Bilingualism* 15(4). 435–453. DOI: [10.1080/13671433.2012.665828](https://doi.org/10.1080/13671433.2012.665828).
- Quay, Suzanne & Sarah Chevalier. 2019. Fostering multilingualism in childhood. In Simona Montanari & Suzanne Quay (eds.), *Multidisciplinary perspectives on multilingualism*, 205–227. Berlin: De Gruyter. DOI: [10.1515/9984-010](https://doi.org/10.1515/9984-010).

Suzanne Quay

- Quay, Suzanne & Simona Montanari. 2016. Early bilingualism: From differentiation to the impact of family language practices. In Elena Nicoladis & Simona Montanari (eds.), *Bilingualism across the lifespan: Factors moderating language proficiency*, 23–42. Washington, D.C. & Berlin: American Psychological Association (APA) & De Gruyter Mouton.
- Quin, Yow Wei, Ferninda Patrycia & Suzanne Flynn. 2016. Code-switching in childhood. In Elena Nicoladis & Simona Montanari (eds.), *Bilingualism across the lifespan: Factors moderating language proficiency*, 81–100. Washington, D.C. & Berlin: American Psychological Association (APA) & De Gruyter Mouton.
- Ruby, Mahera. 2012. The role of a grandmother in maintaining Bangla with her granddaughter in East London. *Journal of Multilingual and Multicultural Development* 33(1). 67–83. DOI: [10.1080/01434632.2011.638075](https://doi.org/10.1080/01434632.2011.638075).
- Sankoff, David & Shana Poplack. 1981. A formal grammar for code-switching. *Research on Language & Social Interaction* 14(1). 3–45. DOI: [10.1080/08351818109370523](https://doi.org/10.1080/08351818109370523).
- Schalley, Andrea C. & Susana A. Eisenclas. 2020. *Handbook of home language maintenance and development: Social and affective factors*. Berlin: De Gruyter Mouton.
- Schalley, Andrea C. & Susana A. Eisenclas. 2022. Parental input in the development of children’s multilingualism. In Anat Stavans & Ulrike Jessner (eds.), *The Cambridge handbook of childhood multilingualism*, 278–303. Cambridge: Cambridge University Press. DOI: [10.1017/69771.016](https://doi.org/10.1017/69771.016).
- Schott, Esther, Lena V. Kremin & Krista Byers-Heinlein. 2021. The youngest bilingual Canadians: Insights from the 2016 census regarding children aged 0–9. *Canadian Public Policy* 48(2). 254–266. DOI: [10.3138/cpp.2021-064](https://doi.org/10.3138/cpp.2021-064).
- Schroedler, Tobias, Judith Purkardhofer & Katja F. Cantone. 2024. The prestige and perceived value of home languages. *Insights from an exploratory study on multilingual speakers’ own perceptions and experiences of linguistic discrimination*. *Journal of Multilingual and Multicultural Development* 45(9). 3762–3779. DOI: [10.1080/01434632.2022.2121402](https://doi.org/10.1080/01434632.2022.2121402).
- Sevinç, Yeşim. 2020. Anxiety as a negative emotion in home language maintenance and development. In Andrea C. Schalley & Susana A. Eisenclas (eds.), *Handbook of home language maintenance and development: Social and affective factors*, 84–108. Berlin: De Gruyter Mouton. DOI: [10.1515/9175-005](https://doi.org/10.1515/9175-005).
- Shen, Chunxuan & Wenying Jiang. 2021. Heritage language maintenance and identity among the second-generation Chinese-Australian children. *Bilingual Research Journal* 44(1). 6–22. DOI: [10.5882.2021.1890650](https://doi.org/10.5882.2021.1890650).
- Silva-Corvalán, Carmen. 2014. *Bilingual language acquisition: Spanish and English in the first six years*. Cambridge: Cambridge University Press.

7 Supporting bilingual development through code-switching

- Slavkov, Nikolay. 2017. Family language policy and school language choice: Pathways to bilingualism and multilingualism in a Canadian context. *International Journal of Multilingualism* 14(4). 378–400. DOI: [10.0718.2016.1229319](https://doi.org/10.0718.2016.1229319).
- Smith-Christmas, Cassie. 2014. Being socialised into language shift: The impact of extended family members on family language policy. *Journal of Multilingual and Multicultural Development* 35(5). 511–526. DOI: [10.1080/01434632.2014.882930](https://doi.org/10.1080/01434632.2014.882930).
- Spolsky, Bernard. 2012. Family language policy – The critical domain. *Journal of Multilingual and Multicultural Development* 33(1). 3–11. DOI: [10.1080/01434632.2011.638072](https://doi.org/10.1080/01434632.2011.638072).
- Stavans, Anat & Ulrike Jessner. 2022. *The Cambridge handbook of childhood multilingualism*. Cambridge: Cambridge University Press.
- Toribio, Almeida Jacqueline. 2004. Convergence as an optimization strategy in bilingual speech: Evidence from code-switching. *Bilingualism: Language and Cognition* 7(2). 165–173. DOI: [10.1017/S4001476](https://doi.org/10.1017/S4001476).
- Tran, Van H., Sarah Verdon & Sharynne McLeod. 2022. Consistent and persistent: Successful home language maintenance among Vietnamese-Australian families. *Journal of Home Language Research* 5(1): 5. 1–19. DOI: [10.16993/jhlr.43](https://doi.org/10.16993/jhlr.43).
- Treffers-Daller, Jeanine. 2022. Code-switching among bilingual and trilingual children. In Anat Stavans & Ulrike Jessner (eds.), *The Cambridge handbook of childhood multilingualism*, 190–214. Cambridge: Cambridge University Press. DOI: [10.1017/69771.011](https://doi.org/10.1017/69771.011).
- Wong, Anita Mei-Yin. 2023. *Understanding development and disorder in Cantonese using language sample analysis*. New York: Routledge.
- Wong Fillmore, Lily. 1991. When learning a second language means losing the first. *Early Childhood Research Quarterly* 6(3). 323–346. DOI: [10.1016/S\(05\)80059-6](https://doi.org/10.1016/S(05)80059-6).
- Yip, Virginia & Stephen Matthews. 2000. Syntactic transfer in a Cantonese-English bilingual child. *Bilingualism: Language and Cognition* 3(3). 193–208. DOI: [10.1017/S136672890000033X](https://doi.org/10.1017/S136672890000033X).
- Yip, Virginia & Stephen Matthews. 2007. *The bilingual child: Early development and language contact*. Cambridge: Cambridge University Press. DOI: [10.1017/CBO9780511620744](https://doi.org/10.1017/CBO9780511620744).
- Yip, Virginia & Stephen Matthews. 2016. Code-mixing and mixed verbs in Cantonese-English bilingual children: Input and innovation. *Languages* 1(1). 4. DOI: [10.3390/languages1010004](https://doi.org/10.3390/languages1010004).
- Yow, Wei Quin, Jessica S. H. Tan & Suzanne Flynn. 2018. Code-switching as a marker of linguistic competence in bilingual children. *Bilingualism: Language and Cognition* 21(5). 1075–1090. DOI: [10.1017/S1366728917000335](https://doi.org/10.1017/S1366728917000335).

Suzanne Quay

Zong, Jie & Jeanne Batalova. 2015. *The limited English proficient population in the United States in 2013*. <https://www.migrationpolicy.org/article/limited-english-proficient-population-united-states-2013>.

Chapter 8

Code-switching by individuals with neurodevelopmental conditions: A scoping review

Draško Kaščelan^a & María del Carmen Parafita Couto^b

^aUniversity of Essex ^bLeiden University

Code-switching or the use of more than one language in an utterance or in conversation is a natural and often quite frequent phenomenon in bilingualism. While there have often been concerns about bilingualism in neurodivergent populations, to date, no comprehensive overview has been done on code-switching practices among neurodivergent individuals (children or adults). In this scoping review, we aim to: (i) identify neurodevelopmental conditions in which code-switching has been investigated; (ii) identify approaches used to explore code-switching in this population; (iii) describe the demographic and bilingualism-related characteristics of investigated populations; (iv) outline any comparisons in code-switching practices between neurodivergent and neurotypical bilinguals; (v) outline attitudes towards code-switching in the investigated population. Among the scarcely conducted research on the topic (n = 31 sources), we uncover that code-switching has primarily been explored in individuals with autism and with language disorder. The review identifies under-representation in research from most areas outside of North America and Europe, as well as under-representation in research with older children and adults. We discuss the plethora of methodological approaches used to explore code-switching in neurodivergent populations and their ecological validity. The characteristics of code-switching in neurodivergent individuals are also addressed, as well as the implications of these findings for educators, speech and language therapists, researchers, and bilingual families with neurodivergent family members.

1 Introduction

In recent years, research on bilingualism in various neurodevelopmental conditions has been expanding.¹ This has been particularly observed in conditions such as autism and Developmental Language Disorder (DLD). A large portion of this work has been motivated by parental or practitioner concerns regarding bilingual development in this population (e.g., [Kay-Raining Bird et al. 2012](#)). Specifically, does acquiring more than one language pose challenges for neurodivergent children? In light of this question, several reviews of the literature have emerged indicating no detrimental effects of bilingualism² either across neurodevelopmental conditions ([Uljarević et al. 2016](#)) or in some conditions such as autism ([Prévost & Tuller 2021](#)). Nevertheless, a variety of methodological confounds and unexplored areas, as noted in [Prévost & Tuller \(2021\)](#), might prevent us from generalising these findings across the population. One aspect of bilingualism which has not been addressed in existing reviews of bilingualism and neurodevelopmental conditions is code-switching. Code-switching or the use of more than one language in an utterance or in conversation is a natural and often quite frequent phenomenon in bilingualism. Taking a clause as a unit of analysis, code-switching can include:

- (1) a. A switch within a clause (intra-clausal switching); a Welsh-English example taken from the Siarad corpus, Stammers6, line 6 ([Deuchar et al. 2014](#))

Mae yn lle lovely
 be.V.3S.PRES in.PREP where.INT love.SV.INF+ADV
 ‘It’s a lovely place’
- b. A switch between clauses (inter-clausal switching); a Spanish-English example taken from the Miami corpus, Herring12, line 165 ([Deuchar et al. 2014](#))

Quiero ir quiero ir but
 want.V.1S.PRES go.V.INF want.V.1S.PRES go.V.INF but.CONJ
 like I don’t wanna
 like.CONJ I.PRON.SUB.1S do.V.1S.PRES+NEG want.V.INF+TO.PREP

¹We note that the introduction and the methods sections of this chapter are expanded and adapted versions of the content written up for the pre-registered protocol of this scoping review (see [Kaščelan & Parafita Couto 2022](#)).

²In this paper we often use the terms bilingual(ism) and multilingual(ism) interchangeably.

8 Code-switching by individuals with neurodevelopmental conditions

go by myself

go.V.INF by.PREP myself.PRON.REFL.1S

‘I want to go, I want to go but, like, I don’t wanna go by myself’

Apart from its prevalence, work on code-switching in neurotypical populations has indicated that code-switching gives a unique insight into the nature of one’s linguistic competence (Poplack 1980, Yow et al. 2018, Truscott & Sharwood Smith 2017), language representation (Truscott & Sharwood Smith 2017), and potential cognitive mechanisms of bilingual language control (Hofweber, Marinis, and Treffers-Daller 2019). Furthermore, code-switching is an important indicator of linguistic community norms (Deuchar 2020, Parafta Couto & Balam forthcoming) which need to be acquired for speakers to be the fully functioning members of a specific group (Phillips & Deuchar 2022). Taking into account the importance of code-switching for understanding bilingual development and lived experiences of bilinguals, it is vital to identify to which extent code-switching has been explored in neurodivergent individuals. Furthermore, considering the role that code-switching plays in the formation of bilingual identities (e.g., Myers-Scotton 1983, Yim & Clément 2021) as well as the importance of maintaining bilingualism for child wellbeing (Müller et al. 2020), it is necessary to offer state-of-the-art guidance for practitioners (such as educators and speech and language therapists) in terms of advice on code-switching to provide to bilingual families with neurodivergent family members.

To address this, we aim to conduct a scoping review of code-switching in bilinguals with various neurodevelopmental conditions. There are several reasons for opting to use the scoping review methodology to achieve this aim rather than rely on alternative forms of evidence synthesis. Specifically, scoping reviews are best suited to explore the breadth of research - that is, they show what evidence exists rather than what is effective (Munn et al. 2022). Furthermore, they are ideally suited for mapping out and summarising the available evidence, identifying gaps in research/knowledge, and informing future research or systematic reviews with more specific research questions (Peters et al. 2020, Tricco et al. 2018). Scoping reviews can also include work implementing quantitative, qualitative or mixed methods, as well as various sources of evidence - published academic research, unpublished work, other reviews, or non-academic documents, such as policy documents or online media (Peters et al. 2021). In line with the aims of the scoping review methodology, we are casting our search net wide in order to identify any research on code-switching in individuals with neurodevelopmental conditions. As it will be discussed further below, this will lead to the use of terminology and identification of related but different language mixing phenomena

that do not necessarily correspond to the definition of code-switching which we provided in examples (1a) and (1b) above. These differences and their implications are highlighted in the discussion section.

In the following section we outline five research objectives of this scoping review addressed through ten research questions. In the methodology section, the procedure implemented for conducting this scoping review is comprehensively described. This is followed by the results presentation and the discussion of findings with implications for the livelihood of neurodivergent bilinguals, as well as for educational, and speech and language therapy practice.

2 Objectives

This scoping review aims to map the available evidence on code-switching in bilinguals with neurodevelopmental conditions across bilingual contexts in order to: (i) identify neurodevelopmental conditions in which code-switching phenomena has (not) been investigated; (ii) identify approaches used to explore code-switching in this population; (iii) describe the demographic and bilingualism-related characteristics of investigated populations; (iv) outline any comparisons in code-switching practices between neurodivergent and neurotypical bilinguals; (v) get a sense of attitudes towards code-switching and advice given regarding code-switching in the investigated population. The following research questions will be addressed:

1. In which neurodevelopmental conditions has code-switching been investigated? [*objective (i)*]
2. Was code-switching the main focus of investigation or addressed sporadically? [*objective (ii)*]
3. Which method was used to analyse the code-switching data: quantitative, qualitative, mixed method? [*objective (ii)*]
4. Which approach was used to explore code-switching in this population: spontaneous speech, semi-experimental approach, experimental approach, a combination? [*objective (ii)*]
5. What was the number of participants in studies investigating code-switching in neurodivergent bilinguals? [*objective (iii)*]

8 Code-switching by individuals with neurodevelopmental conditions

6. What were the demographic characteristics of communities in which code-switching has been investigated (geographic location, language combinations, age range of participants, participants' sex and gender distribution, socioeconomic status)? [*objective (iii)*]
7. Have studies of code-switching in neurodivergent individuals included a neurotypical bilingual comparison group? [*objective (iv)*]
8. Are there any indications that code-switching happens differently in neurodivergent individuals in comparison to neurotypical ones? [*objective (iv)*]
9. What kind of professional advice have respective individuals received regarding code-switching: positive, negative, both, neither, no data? [*objective (v)*]
10. How is code-switching perceived in the environment of neurodivergent individuals: positively, negatively, both, neither, no data? [*objective (v)*]

3 Methods

3.1 Protocol and pre-registration

A scoping review protocol was prepared and registered prior to conducting the review in accordance with the Joanna Briggs Institute (JBI) methodology for scoping reviews and informed by the guidance provided in [Peters et al. \(2022\)](#) and [Pollock et al. \(2022\)](#). The registration of the protocol was conducted on the Open Science Framework (OSF) on 17 August 2022 (<https://osf.io/ru8pe>). Throughout the chapter, we specify in footnotes whether there were any departures from the pre-registered protocol and clarify the rationale for those decisions. At the time of pre-registering the protocol, we did not identify any previous or ongoing reviews on the topic in The International Prospective Register of Systematic Reviews (PROSPERO) database. The PROSPERO database has also been checked recently, and we identified a protocol for a systematic review on code-switching by [Snijders et al. \(2023\)](#) registered on 03 December 2023. However, as the registered systematic review protocol by [Snijders et al. \(2023\)](#) focuses on code-switching in neurotypical children between the ages of 2 and 6 years, it does not address the aims proposed in this scoping review.

3.2 Eligibility criteria and information sources

We strictly adhered to the decisions on eligibility criteria and information sources as reported in the pre-registered protocol (Kaščelan & Parafita Couto 2022). The table and the text in this section are close adaptations from the protocol. Table 1 reports the eligibility criteria for inclusion through a summary of the population, the concept, and the context of interest.

Table 1: Specification of the Population, Concept, Context (PCC) of the scoping review.

PCC Element	Specification
Population	We focused on bilingual individuals (of any age, gender, sex, socioeconomic status) with any neurodevelopmental conditions as defined by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; American Psychiatric Association 2013) with few additions of terms of our own as explained in the search strategy below. We included papers even if they simply stated that participants had a particular diagnosis, without a need for a formal confirmation of how the diagnosis had been made. Research with participants with more than one condition was also included.
Concept	The concept under investigation is code-switching. In this review, we use <i>code-switching</i> as an overarching term for any language mixing practices (i.e., use of more than one language in an utterance or in conversation). All types of code-switching were eligible for inclusion. See the search strategy below for the specification of code-switching types and terms which we used when searching the literature.
Context	The scoping review focuses on any bilingual/multilingual communities with no restrictions regarding the geographic location or the nature of bilingualism (e.g., simultaneous, sequential, etc.). See the search strategy below for the specification of bilingualism-related terms which were used when searching the literature.

There were no restrictions for inclusion in terms of methodology as we aimed to include quantitative, qualitative, and mixed methods studies. Furthermore, any

8 *Code-switching by individuals with neurodevelopmental conditions*

approach to investigating code-switching was included: work with spontaneous speech, semi-experimental approach, experimental approach, a combination of these.

The scoping review includes published articles, chapters, books, and conference proceedings, as well as unpublished and grey literature, such as preprints and dissertations. As indicated in [Kaščelan & Parafita Couto \(2022\)](#), work excluded at the stage of full text screening can be found via a link provided in supplementary materials, which leads to an OSF page containing a complete dataset and the exclusion reasons specified.

Whenever we identified any sort of evidence synthesis on this topic, rather than including it in the review, we followed up on the original papers reviewed in the identified synthesis. Nevertheless, the identified synthesis papers are still acknowledged in our dataset (see supplementary materials) as sources of evidence.

Identification of resources was conducted through the search of databases and through alternative ways of identifying relevant work. The following electronic databases were used as a source of information: Ovid (Embase Classic+Embase, Global Health, APA PsycInfo), ProQuest, Scopus, and Web of Science.³ The databases were searched on 13 March 2023. It was decided not to have date restrictions for any database, as we aimed to include any relevant work in our review. The start and the end date varied across databases due to the scope of time that each database has. The project OSF page (<https://osf.io/4nwgd/>) contains an example report of a search conducted in one of the databases and it illustrates the date restrictions imposed by the database scope (e.g., for APA PsychInfo, this was from 1806 to February Week 4 2023).

In addition to the database search, there were three alternative ways of identifying relevant work. First, prior to conducting the database search, we identified some relevant work through our knowledge of and interaction with the literature on bilingualism and neurodevelopmental conditions. Second, in order to obtain published and unpublished work which might not be picked up in the database search, we sent a notice about the scoping review (see a draft notice in the protocol Appendix) via several mailing lists and networks (ISB_list, Info-CHILDES, Bi-SLI COST Action, The LINGUIST List) and on social media (Twitter, Facebook). In this notice, we asked colleagues to send us any (up)published work on code-switching in bilinguals with neurodevelopmental conditions if they would

³During the time of this scoping review's protocol drafting, the first author was affiliated with the University of Leeds. Consequently, the University of Leeds library resources were used to select these databases for the final search. These databases were listed as popular for research in health, medicine, psychology, arts, social sciences, and related subjects.

like for it to be included in the review. These notices were sent out in March 2023. Third, after the database search was conducted and throughout the screening process, we monitored relevant literature and included any work which has not already been identified - this work did not go through the abstract screening stage; rather these papers were fully screened. The routes of information sources are summarised in Figure 1, in the identification section.

There were no language restrictions in our search. However, texts in languages not spoken/understood by the authors were excluded. Information on this can be found in the dataset on the project OSF page (<https://osf.io/4nwgd>).

3.3 Search and the selection of sources of evidence

The complete list of key terms used for search in electronic databases can be found in Table 2 of the pre-registered protocol (Kaščelan & Parafita Couto 2022). There were three sets of key terms: (1) Key Term 1 was participant-related (e.g., neurodevelopmental disorder*, autism spectrum condition*, DLD, vocal tic disorder*, etc.), (2) Key Term 2 was concept-related (e.g., code-switch*, alternation*, intra-clausal*, borrow*, translanguag*, etc.), and (3) Key Term 3 was context-related (e.g., bilingual*, BFLA, L2, societal language*, etc.). The list of terms was finalised during several discussions between the two authors. For the purposes of this review, we use the term code-switching interchangeably with the terms such as language mixing or code mixing. Specifically, due to the nature of scoping reviews (which require casting the search net as wide as possible), for the Key Term 2 (concept-related), we use various code-switching terminology which does not necessarily refer to the same phenomena (e.g., inter-clausal, inter-sentential, insertions, translanguaging, language mixing, code-switching, code mixing, etc.). However, we acknowledge the lack of consensus in the field in terms of differentiating between code-switching, language mixing, or code-mixing (see for example Gullberg et al. (2009) for a discussion) as well as different assumptions underlying code-switching and translanguaging (Balam 2021, Li 2018) We return to this point in the discussion of the review findings.

When searching the electronic databases, each term within the category *Key Term 1* was linked with a boolean operator *OR*. The same was done for terms within the categories *Key Term 2* and *Key Term 3*. Following this, the key terms were connected with a boolean operator *AND*, as follows: (*Key Term 1*) *AND* (*Key Term 2*) *AND* (*Key Term 3*). The search strategy was adapted for each included database. An example report of a search conducted in one of the databases, can be found on the project OSF page, including necessary filters and search syntax

8 Code-switching by individuals with neurodevelopmental conditions

adaptations: <https://osf.io/4nwgld/>. The first author conducted the search in all databases following the university library resources for conducting reviews.

The selection of resources is visually represented in Figure 1. We collected 273 resources (i) through the identification of resources by the authors prior to the database search, (ii) through the search of four databases, and (iii) through communication with colleagues in the field. These were exported in an excel file (see the *inclusion-exclusion sheet* in the dataset on the project OSF page). Following this, 121 duplicates were identified and removed. The remaining 152 abstracts were screened independently by both authors and the following exclusion criteria were applied: (1) not available, (2) not in a language that reviewers speak/understand, (3) not about bi/tri/multilingualism, (4) not about code-switching, (5) not about neurodevelopmental conditions, or (6) other (specification required). Independent abstract screening resulted in disagreement regarding 13 out of 152 abstracts (i.e., in only 8.6% of cases). The authors discussed the disagreements until a consensus was reached.

This left us with 34 papers for full text screening. A randomly selected subset of these (approximately a third: 11/34 or 32.4%) was independently screened fully by both authors.⁴ The authors disagreed in 4 out of 11 cases (i.e., in 36.4% of cases). Considering the large exclusion/inclusion disagreement rate in full text screening, the authors had an extensive discussion until consensus was reached. Following that discussion, the first author fully screened the rest of the data, consulting the second author whenever in doubt. This left us with 14 papers for inclusion in the review.

Following this, the reference list of the 34 papers that were fully screened were also checked for any relevant literature.⁵ This led us to identifying 25 references which could potentially be included in the review. The full text screening of these 25 papers was completed by the first author in consultation with the second author whenever in doubt. This resulted in: excluding 1 resource due to its inaccessibility, excluding 18 resources for additional reasons as per the previously outlined exclusion criteria, and keeping 6 resources in the review. The

⁴This was one of the rare cases in which we departed from the pre-registered protocol. According to the protocol, we intended to independently complete full text screening of all the resources. However, due to personal circumstances of the second author and time restrictions, the second author fully screened a randomly selected subset (i.e., approximately a third of the dataset).

⁵This is another departure from the pre-registered protocol. Specifically, according to the protocol, we aimed to screen the references only in the papers included in the review after the full text screening, which in this case would be 14 papers. However, to avoid missing relevant work, we screened the references of all 34 papers that were fully screened, thus increasing the chances of identifying relevant work.

Draško Kaščelan & María del Carmen Parafita Couto

lists of these 25 papers identified in references with reasons for exclusion and additional notes can be found in the *identified in references* sheet in the datafile on the project OSF page.

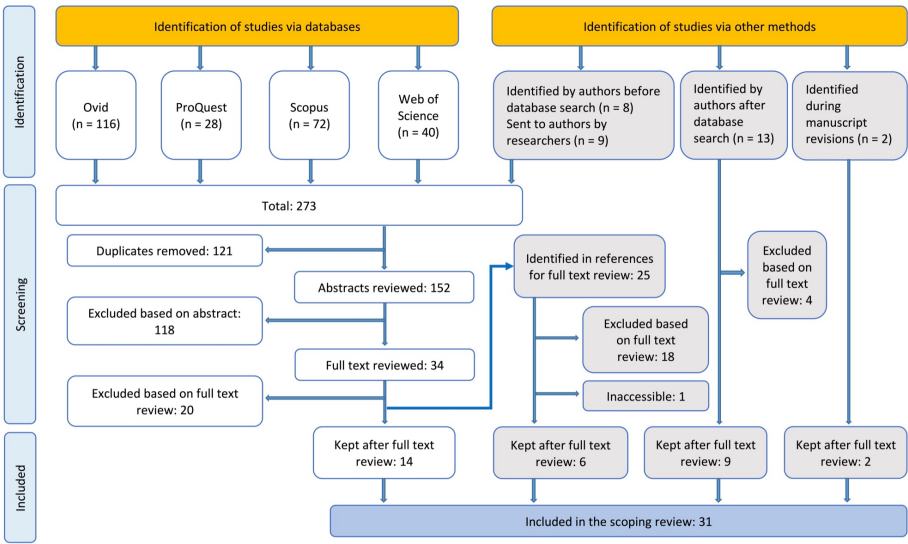


Figure 1: The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) flowchart of the scoping review process.

While we were screening the resources identified in the databases and gathered through the above outlined alternative approaches, we also monitored relevant literature for any new research published after our database search or for any research that we potentially missed. Over the period of about 9 months, we identified 13 potential resources, which were fully screened by the first author in consultation with the second author when required. Of these, 4 resources were excluded and 9 were kept for inclusion in the scoping review. The details about these papers, the exclusion reasons and relevant notes can be found in the *identified after review* sheet in the datafile on the project OSF page.

Finally, during this chapter’s revision process, we were sent a recently published paper on the topic by the paper authors (Blom et al. 2024) to consider for inclusion in the scoping review. In addition, in Blom et al.’s (2024) references, we identified another relevant paper for inclusion (Gross & Castilla-Earls 2023).

8 Code-switching by individuals with neurodevelopmental conditions

Both of these papers were included in the scoping review (the details about these papers can be found in the *identified during revisions* sheet in the datafile on the project OSF page). Since they were both published after our database search (i.e., after 13 March 2023), it is clear why they were not picked up via the key term search.

Following this procedure, we were left with 31 papers for data extraction and analysis. From these 31 papers, 27 were published / accepted for publication (22 journal articles, 4 book chapters, 1 proceeding) and 4 were unpublished dissertations.⁶ The above outlined process is visually represented in Figure 1.

3.4 Data charting process and data items

A data charting form was developed jointly by both authors. The developed version was posted on the project OSF page with the pre-registered protocol on 17 August 2022 (sheet *data extraction sheet*). Of the 31 resources for data extraction, 3 were randomly selected (Gutiérrez-Clellen et al. 2009, Iluz-Cohen & Walters 2012, Soesman et al. 2022) and their data was independently extracted by both authors in the existing data charting form. Following this, the authors met to discuss the adequacy of the data charting form and made minor changes. A column titled *Gender or Sex explicitly labelled in the paper* was added for specification when resources did not distinguish between sex and gender and when we had to assume which one they referred to in their descriptive statistics. The column *Group differences in code-switching* (y/n) was renamed to *Group differences/characteristics of code-switching* to allow the extraction of data when a comparison group was not included but rather only the description of code-switching practices of neurodivergent individuals were available. The column *Group differences in code-switching code* was added to allow for a coding of whether the groups (when present) differ in their code-switching practices and to what extent: no, both (predominantly no), both, both (predominantly yes), yes, inconclusive, or not applicable (na). We also added a column *Comments* for any additional points or notes that might facilitate data analysis. Following this data charting form calibration, the first author extracted the data from the rest of the resources. Once this step was completed, the authors met to discuss the data and resolve any difficulties in data extraction or classification, particularly in relation to the decisions on the degree of differences in code-switching identified between neurodivergent and neurotypical individuals or the lack of equivalency in data on participants'

⁶As pointed out by an anonymous reviewer, a part of the dissertation by Angulo Jiménez (2020) has recently been published as a journal article (Angulo-Jiménez & DeThorne 2024). Consequently, we acknowledge this publication here and in the reference list.

socioeconomic status. Throughout the data extraction and analysis process, few other columns were added to the data extraction sheet as agreed by both authors. The datafile (sheet *data extraction sheet*) contains a note on these columns indicating that they were added after the protocol registration. The complete datafile can be accessed on the project OSF page (see Supplementary materials for the link to the OSF page).

We extracted the data on the following: resource description (document number, author(s), publication year, title, document type), objective I (neurodevelopmental condition(s) investigated), objective II (whether code-switching was the main focus of investigation, the method used to explore code-switching, and the approach used to investigate code-switching), objective III (number of neurodivergent participants, geographic location of the studies, language combinations, participant age range, participant sex and gender distribution, participant socioeconomic status), objective IV (inclusion of the neurotypical comparison group, group differences/characteristics of code-switching), and objective V (professional advice on code-switching and perception of code-switching in the environment of neurodivergent individuals).

3.5 Critical appraisal and synthesis of results

Our research objectives did not include explicit critical appraisal of each study included in the review. Consequently, such appraisal was not explicitly conducted. However, in the results and the discussion sections, the analysis of the extracted data will include the quality assessment of gathered evidence. In particular, we will address the contributions and limitations of existing research as well as its implications for future research and the effects that existing approaches might have on the livelihood of the target community members, educational and speech and language therapy practice.

The resource identification, screening, eligibility, and inclusion are outlined in Figure 1. The datafile on the project OSF page includes a list of all the studies that went through the outlined process and reasons for exclusion with additional notes where relevant. In the sections below, we will use a combination of tables, figures and narrative to present the data and address all five objectives of the scoping review.

4 Results

4.1 Scope of investigations and focus

Despite the long list of neurodevelopmental conditions considered in this review, code-switching has been investigated only in a limited number of conditions as indicated in Table 2. Specifically, language disorder and autism seem to be prevalently examined in relation to language mixing, while other neurodevelopmental conditions tend to appear sporadically or often as secondary conditions in the investigated samples.

Among the 31 studies, code-switching was the main focus in 18 of them (58%). It is important to note that out of 13 studies which explored code-switching sporadically, only 1 was identified via the database search, which is expected as the key terms targeted the studies focusing on all aspects of PCC (see Table 1 above). From the remaining 12 studies, 4 were identified in references of screened papers, and 8 were identified after the database search. All of these 8 studies were picked up due to the first author's familiarity with the literature on autism and bilingualism. Consequently, this approach inflated the number of studies focusing on autism indicated in Table 2.⁷ However, even if these studies were excluded, autism would still represent a commonly investigated condition in the corpus that was identified.

4.2 Methods and approaches to code-switching

Quantitative method was used predominantly across the studies ($n = 20$), followed by the qualitative method ($n = 10$), while the mixed method was used only in one study (Angulo Jiménez 2020). A series of approaches taken to explore code-switching can be broadly grouped as follows: spontaneous speech ($n = 6$), semi-experimental ($n = 5$), experimental ($n = 6$), a combination of these ($n = 4$), not applicable ($n = 10$). The studies implementing the spontaneous speech approach included guided conversations with obligatory questions about family, school, games (Sanz-Torrent et al. 2007), mundane naturalistic conversations (Angulo Jiménez 2020, Yu 2013), typical daily interactions (Cohen et al. 2025), interactions with a therapist (Ponce-Lawler 2017), and a written diary (Karniol 1992). Studies

⁷The percentages in the table do not add up to 100% as some studies had participants with more than one condition. In this table, the term language disorder is including terms such as language impairment, specific language impairment, developmental language disorder, primary language impairment. The term autism includes terms such as autistic disorders, high functioning autism, pervasive developmental disorder, pervasive developmental disorder – not otherwise specified, Asperger's syndrome, autism spectrum disorder.

Table 2: Frequency of explored neurodevelopmental conditions.

Neurodevelopmental condition	Number of studies	% of studies
Language disorder	16	52%
Autism	14	45%
ADHD	1	3%
Down Syndrome	1	3%
Non-verbal learning disorder	1	3%
Prader-Willi Syndrome	1	3%
Repaired cleft palate	1	3%
Stuttering	1	3%

with semi-experimental designs relied on story retellings (Aguilar-Mediavilla et al. 2024, Kapantzoglou et al. 2021), narrative elicitation (Mammolito 2015), or narrative elicitations and story retellings (Gross & Castilla-Earls 2023, Iluz-Cohen & Walters 2012). The experimental studies included a range of tasks: cued picture selection (King et al. 2021), cued picture naming (Snijders et al. 2025), sentence repetition (Soesman et al. 2022), acceptability judgement and sentence completion (Liceras & Garcia-Alcaraz 2019), repeated word association (Sheng et al. 2012), and expressive language assessments (Pert 2007). The four studies which combined these approaches included an experimental and a semi-experimental task (guessing and description phases of a scripted confederate dialogue task, Gross & Kaushanskaya 2022), or semi-experimental tasks (narrative/story retell and/or elicitation) and a spontaneous conversation (Blom et al. 2024, Gutiérrez-Clellen et al. 2009, Restrepo 2003). Ten studies did not include any particular approach to code-switching as these were among the studies in which code-switching was not the main focus of research. We return to the implications of using various approaches to examine code-switching in the discussion section.

4.3 Number of neurodivergent bilinguals and demographics

Across 31 identified papers, there were in total 373 neurodivergent (range: 1-49) and 401 neurotypical (range: 3-65) participants. The 31 papers included 32 studies in total. Focusing on the neurodivergent participants only, about a third of the studies (n = 11) could be classified as case studies (1-3 participants), while 13 studies had between 4 and 20 participants. Only 6 studies included data about more than 20 participants (range: 23-49). Finally, 2 studies included research with practitioners as participants rather than neurodivergent individuals.

8 *Code-switching by individuals with neurodevelopmental conditions*

In terms of geographic distribution reported in the identified 31 papers, the neurodivergent participants resided predominantly in the USA ($n = 17$ studies), across various parts of the country (as specified in some of the studies): a mid-western city, a metropolitan area in north-western USA, southern California, metro Atlanta area, El Pueblo - Midwest, El Paso - Texas, Greater San Francisco Bay area, Austin - Texas, Denver - Colorado, Minneapolis, along the USA-Mexican border, California, Massachusetts, Boston, Western New York and South Texas. Canada, Israel, and the UK (England, Wales, Rochdale - north of England) were each represented by three studies, while there were two studies from France, two from Spain (Catalonia, Balearic Islands, Mallorca), and two from the Netherlands. The following countries were each represented in one study only: Greece, Egypt, Ireland, and Singapore. The visual representation of countries included in the studies identified by the scoping review can be seen in Figure 2.

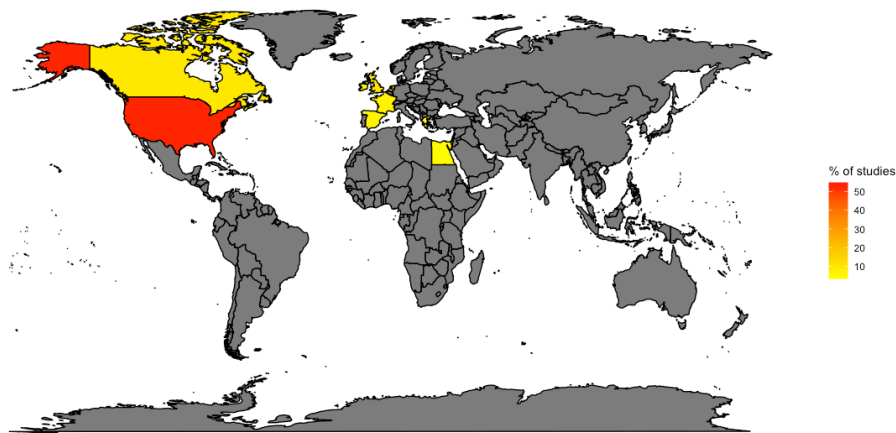


Figure 2: Percentage of studies on code-switching in neurodevelopmental conditions per country.

Considering the language combinations of neurodivergent participants, more than a third of the studies included English-Spanish bilinguals ($n = 12$), while 7 studies included a mix of participants speaking English and another language (Arabic, Armenian, Bangala, Danish, French, Gujarati, Hebrew, Hindi, Italian, Katch, Lithuanian, Mandarin, Polish, Portuguese, Punjabi, Russian, Spanish, Turkish, Urdu, Welsh). The following language combinations were each represented in two studies: Hebrew-English, English-Mandarin, Catalan-Spanish, Dutch-Turkish. Finally, there was 1 study focusing on each of the following

language combinations: English-Irish, English-Mirpuri-Urdu, Hebrew-English-Hungarian, French and another language (English, Gujarati, Hindi, Marathi, Punjabi, Spanish). Looking at these findings, all but four studies (Aguilar-Mediavilla et al. 2024, Blom et al. 2024, Sanz-Torrent et al. 2007, Snijders et al. 2025) focused on participants speaking English in combination with another language.

The majority of studies (23/31) were conducted with children, while only 2 focused on adults and 2 included data from both children and adults. Data about age was missing or irrelevant in 4 studies. As shown in Figure 3, most studies have been conducted with pre-school and early years school children. Gender or sex were explicitly labelled in 12 studies, only 1 of which (Angulo Jiménez 2020) inquired about both sex and gender. In the remaining 19 studies, the distinction was not explicitly mentioned, and 5 of these studies included no data on neurodivergent participants' gender/sex. For the purposes of this review's summary, all the binary sex/gender data is collated into male and female. Across the 31 studies, there were 199 male and 68 female neurodivergent participants. Two studies did not provide gender/sex breakdown of participants within the neurodivergent group; therefore, these participants (23 neurodivergent children in King et al. (2021), and 13 neurodivergent children in Soesman et al. (2022)) are not included in our count.

The final demographic variable which we summarise is the socioeconomic status (SES) of participants. The breakdown across the studies was as follows: 7 studies included participants with low SES, 1 with medium SES, 5 with high SES, 3 with mixed SES, and 15 with no data on SES. Despite the authors' consensus in coding the SES data, some of our decisions in coding might have misrepresented the SES status of the participants. For instance, if a study included a mean/median value of the SES status and a large range of scores without the clear indication of how the majority of the sample could be characterised, we labelled this as mixed SES (e.g., in the study by Snijders et al. (2025)). This however obscures the accurate representation of the sample. Furthermore, if the education of parents was the only SES indicator and the parents had a university degree of any level, we would code this as high SES (e.g., the study by Yu (2016)). Nevertheless, this approach is not necessarily an accurate representation, since families with a university degree do not necessarily have a high annual income.

As expected, SES was operationalised differently across the studies, thus making the task of data comparability quite challenging. About 42% of the studies included education as an SES estimate ($n = 13$). Other indicators included: annual income ($n = 3$), occupation ($n = 3$), free/reduced lunch ($n = 3$), employment ($n = 1$), class indication ($n = 1$), low-high label ($n = 1$), missing data ($n = 15$). The

8 Code-switching by individuals with neurodevelopmental conditions

raw numbers do not add up to the total number of papers included in this review (n = 31), as some of them included more than one SES indicator. A general summary of all the demographic variables can be found in Figure 3.

Gender/Sex		N	Sample size		N	
Female		68	1-3	11		
Male		199	4-20	12		
No data		36	Over 20	6		
Age (in years)		N	Not applicable		2	
4 or under		3	Country		N	
5 or under		3	Canada	3		
6 or under		6	Egypt	1		
7 or under		5	France	2		
8 or under		3	Greece	1		
9 or under		1	Ireland	1		
11 or under		1	Israel	3		
12 or under		1	Netherlands	2		
21 or under		1	Singapore	1		
22 or under		1	Spain	2		
26 or under		1	UK	3		
34 or under		1	USA	17		
No data / Not applicable		4	Language combinations		N	
Socioeconomic status operationalisation		N	Socioeconomic status level		N	
			English-Spanish		12	
			English + another language		7	
Education		13	Low	Hebrew-English		2
Annual income		3		English-Mandarin		2
Free/reduced lunch		3		Catalan-Spanish		2
Occupation		3	Medium	Dutch-Turkish		2
Employment		1		English-Irish		1
Class indication		1		English-Mirpuri-Urdu		1
Low-high label		1	High	Hebrew-English-Hungarian		1
No data		15		French + another language		1
			Mixed			
			No data			

Figure 3: Demographic data of neurodivergent participants.

4.4 Characteristics of neurodivergent code-switching and comparison groups

One of the main goals of this review was to outline characteristics and any possible differences in code-switching practices (e.g., frequencies of code-switching, types of code-switching, etc.) of neurodivergent individuals in comparison to neurotypical code-switching. Out of 31 papers, 14 included a comparison group / individuals, 16 included no comparison group/individuals, and in case of 1 study, comparison was not applicable as it explored professionals' perceptions on language impairments in bilinguals (O'Toole & Hickey 2013). Four of the 16 studies which included no comparison groups/individuals, compared the code-switching patterns of their neurodivergent participants to existing literature on neurotypical code-switching. This data is summarised in Table 3.

Considering the different methodological approaches used across the studies as well as different research aims, direct comparability of findings poses a significant challenge. Below we summarise the findings in four categories: (1) studies that identified no or predominantly no differences between neurodivergent and neurotypical code-switching ($n = 9$), (2) studies that identified different or predominantly different patterns of code-switching by neurodivergent individuals ($n = 5$), (3) studies which identified both differences and similarities between the neurodivergent and neurotypical groups ($n = 3$), and (4) studies with inconclusive results ($n = 1$).

The 9 studies which included no (or predominantly no) differences in code-switching between neurodivergent and neurotypical participants included a variety of tasks/approaches. Examples of no differences include the following: (a) no differences in the frequency of code-switching on language assessments, on a story (re)telling task, on a narrative elicitation task, or in spontaneous speech / a guided conversation (Aguilar-Mediavilla et al. 2024, Gutiérrez-Clellen et al. 2009, Kapantzoglou et al. 2021, Pert 2007, Ponce-Lawler 2017, Sanz-Torrent et al. 2007) (b) no differences in the types of switches, such as the part of speech that is being code-switched, the use of participant- and discourse-related insertions and alternations, or the types of switches used for specific communicative/pragmatic purposes (Angulo Jiménez 2020, Gutiérrez-Clellen et al. 2009, Ponce-Lawler 2017, Yu 2016) (c) no differences in terms of abiding to the Matrix Language Framework constraints in children with SLI (Gutiérrez-Clellen et al. 2009); (d) no differences in gender agreement strategies in an acceptability judgement and a sentence completion task by an adult with Prader-Willi syndrome (Licerias & Garcia-Alcaraz 2019); (e) no differences in the language of insertions, production of words in the phonology of the other language, grammatical violations, and regularisation

8 Code-switching by individuals with neurodevelopmental conditions

Table 3: Comparison of neurodivergent to neurotypical code-switching (data represents number of studies).

Comparison of neurodivergent code-switching to neurotypical patterns	Studies with comparison groups/individuals (n = 14)	Studies with no comparison groups/individuals (n = 16)	Studies in which comparison was not applicable (n = 1)	Total number per difference category
Not different	2	3	0	5
Predominantly not different	3	1	0	4
Both different and similar	3	0	0	3
Predominantly different	2	0	0	2
Different	3	0	0	3
Inconclusive	1	0	0	1
Not applicable	0	12	1	13

errors by children with SLI (Sanz-Torrent et al. 2007). In the four studies which identified predominantly no differences between neurodivergent and neurotypical children, minor differences that were identified included the following: (a) children with SLI producing more verbs in the other language and fewer nouns unlike the age-matched comparison group (Sanz-Torrent et al. 2007); (b) on a story retell task children with DLD showed a higher use of alternations than their age-matched comparison group at the age 8, but this difference disappeared at age 12 (Aguilar-Mediavilla et al. 2024); (c) a child with autism did not code-switch in cases of direct/indirect quotations or to achieve narrative frame breaks (Yu 2016); (d) children with SLI omitted or made errors in late system morphemes on verbs (Pert 2007).

There were 5 studies which identified different patterns of neurodivergent code-switching in comparison to neurotypical expectations. Iluz-Cohen & Walters (2012) identified that in narrative elicitation and story retell tasks children with language impairment code-switched more than neurotypical children both overall and in longer segments of speech. They further identified that unlike neurotypical children, children with language impairment code-switched somewhat more to their first language. On a sentence repetition task using monolingual and code-switched sentences, (Soesman et al. 2022) identified that on English stimuli, children with DLD produced more non-target switches than neurotypical children (non-target switches were instances when a participant produces a code-switch in a sentence repetition task for an item that was originally not code-switched). Furthermore, while the neurotypical children produced more non-target code-switches in Hebrew than in English stimuli, in the DLD group, there were no differences between languages. In a repeated word association task, Sheng et al. (2012) identified that children with language impairment code-switched more from English to Spanish than from Spanish to English, while the neurotypical group demonstrated the opposite pattern. A study by Cohen et al. (2025) compared spontaneous speech of a child with autism to that of his sister and parents. While they identified different frequency of code-switching between family members during various activities and for specific purposes, they demonstrated how integral code-switching can be in multilingual families. Finally, Blom et al. (2024) identified that children with DLD mixed more often in the home language setting than the neurotypical children. Specifically, children with DLD tend to respond in the societal language more than neurotypical children when they are spoken to in the home language. Additionally, societal language proficiency had a stronger impact on mixing by children with DLD than by the neurotypical children. However, Blom et al. (2024) indicated that these differences diminished over time. We note that from these 5 studies which identified differences in neurodivergent code-switching in comparison to neurotypical patterns, only one was based on the analysis of spontaneous speech and it included a case study of an autistic child (Cohen et al. 2025), while another study combined spontaneous speech data with semi-experimental data in their analysis (Blom et al. 2024). We address this point in light of implications for practitioners in the discussions section.

As indicated in Table 3, 3 studies identified both differences and similarities between neurotypical and neurodivergent code-switching. In terms of the identified differences, Gross & Kaushanskaya (2022) found that on a scripted confederate dialogue task, children with DLD were more likely than neurotypical children to produce a response entirely in English when interacting with a monolingual

8 *Code-switching by individuals with neurodevelopmental conditions*

Spanish speaker, even when their proficiency in both languages is taken into account. However, as noted by Gross & Kaushanskaya (2022), cognitive control also played a role in these types of switches and a larger sample would be required to investigate relevant interactions between different factors relevant for children's code-switching practices. In Snijders et al. (2025), on a cued picture naming task, children with DLD made more errors (e.g., cross-language errors, wrong-word errors, and 'I don't know' responses) and had longer reaction times than neurotypical participants. In Gross & Castilla-Earls (2023), in the narrative samples, children with DLD engaged more in between-utterance code-switching than the neurotypical children. Furthermore, Spanish proficiency had a larger negative association with between-utterance switches in the sample of children with DLD. Some qualitative differences between the groups were also identified in terms of types of insertions. We note that these three studies used experimental and/or semi-experimental paradigms, which we address in the discussion section.

Finally, a study by King et al. (2021) produced inconclusive results. Specifically, on a cued language switching task on most trials, children with language impairment had longer reaction times than children without language impairment. However, as the authors acknowledge, the fact that children with language impairment were significantly younger than neurotypical comparisons and that this difference was not accounted for in the analyses suggests that the findings should be interpreted with caution.

4.5 Professional advice and attitudes to code-switching

The majority of reviewed studies did not explicitly inquire about attitudes towards code-switching and professional advice received about the language mixing practices. Specifically, 26/31 studies included no data on professional advice on code-switching, 4 studies included negative advice and only 1 study included both positive and negative advice. Examples of negative advice were as follows:

Parents were also told that they were “creating chaos” in their child's mind by switching between languages and exposing their child to various words. The teachers and therapists then went on to issue further warnings, such as “your child will become terribly confused,” or “become more lost.” This advice was given to families despite the reality that members of the families often spoke no English and would be unable to converse with the children if language choice was restricted. (Jegatheesan 2011: 195)⁸

⁸We note that in some studies, such as in (Jegatheesan 2011), multilingual practice seems to be the norm in the studied families and a practice that they do not want to abandon despite the advice by professionals to stick only to the societal language.

Five of the parents believed that code-switching would cause confusion. For example, Jin said, “In the last 6 months, I’ve tried to speak to Henry as much as possible in English. I would try to speak just in English because people told me not to switch back and forth.” [...] Many of the professionals they encountered, however, appeared to be misinformed about bilingualism and its relationship to the development and learning of children with ASD. Their advice to parents often perpetuated deficit views of bilingualism, including notions that learning two languages would cause semilingualism and delay, that code-switching caused confusion, or that the use of home languages interfered with the learning of English. (Yu 2013: 18-21)

[Jason’s m]other: Learning two totally different languages at the same time and understand them properly and use them properly at the same time, she [the clinician] mentioned about that would be extremely difficult with the kids who have speech problem. (Kremer-Sadlik 2005: 1231)

[O]ne mother (Arabic-speaking father and grandparents) remembered, “I really felt that we were to blame [for his speech delay]. All that switching back and forth caused the problem, not being consistent, and it was all our fault... that we really confused him. I believed he had a language processing disorder, and I believe we hurt him doing that, all that switching.” (Fernandez y Garcia et al. 2012: 12)

Only 1 study (Howard et al. 2021) reported both positive and negative advice on code-switching.

Negative:

We were advised to stick to one language because sometimes it can be very confusing jumping from one language to another, and just to keep that consistency as well. (Nabani). [...] Nabani intimates that she agrees with the advice. (Howard et al. 2021: 185)

Positive:

Eleanora describes this process: ‘we did ask the health visitor when he was little and he just said, “No speak both languages and he’ll be fine”, and that’s what we did and his first words were a mixture, some in English, some in Italian.’ (Howard et al. 2021: 185)

8 *Code-switching by individuals with neurodevelopmental conditions*

Considering attitudes to code-switching, 22/31 studies included no data on this particular matter. From the remaining 9 studies, negative attitudes were present in 5 of them (Howard et al. 2021, Kay-Raining Bird et al. 2012, O'Toole & Hickey 2013, Fernandez y Garcia et al. 2012, Yu 2013), positive attitudes in 1 of them (Oudet et al. 2022), and both positive and negative attitudes in 3 of them (Howard et al. 2021, Kremer-Sadlik 2005, Wharton et al. 2000). An example of each is offered below.

Positive:

The ability to switch between languages in conversation was a welcome development for parents who were concerned about their choice to use their respective L1s with their children. [...] ClaraF2 commented [that] her son's communication progress may have come about through natural developmental maturity. She also spoke about increased time at home during lockdown with both parents speaking their respective L1s in conversation as leading to her son's expressive code-switching. (Oudet et al. 2022: 8)

Negative:

[T]he therapists appeared cognisant of [the] threat [of code-switching in case of endangered languages such as Irish] in noting that the level of code-switching in the bilingual children with SLI was higher than expected. SLT 4: I have come across children who would use a lot more loan words than you would expect them to, so sticking -áil [Irish progressive marker] onto the end of English verbs, that kind of thing. (O'Toole & Hickey 2013: 102)

Both:

Roberta described each language as a 'whole universe' and suggested that code-switching may increase her child's cognitive flexibility: One of the issues with autistic kids is, you know, that they can find it difficult to be flexible in situations, so the fact that he has to switch codes, so with the codes comes a whole universe almost, I think that actually is a good way of practising, you know, flexibility. (Howard et al. 2021: 184)

Anna suggested that the increased challenge of switching between languages may encourage her son Dean to 'keep his mind busy' and avoid distraction. (Howard et al. 2021: 184)

Conversely, some parents felt that the severity of their own child's autistic symptoms—that is the extent of their communicative difficulties—rendered bilingualism unfeasible. After seeing her son distressed by her code-switching, Hira opted for a monolingual approach: Slowly I started working with him at mix-matching and he used to cry and then I said, “it’s fine” and I used to let him cry. “OK, you’re crying, it’s fine”... and we decided just one language. (Howard et al. 2021: 186)

We return to addressing the low frequency of work on this topic and predominantly negative advice and attitudes towards code-switching in the discussion section.

5 Discussion

This scoping review aimed to summarise the academic literature on code-switching in neurodevelopmental conditions. We now turn to discussing the findings of the review aims and drawing implications for researchers, practitioners, and multilingual families. Throughout the discussion, we also acknowledge limitations of the present review.

5.1 Bias in focus

We found that the most frequently studied neurodevelopmental conditions in relation to code-switching are language impairment / DLD and autism. Code-switching is linked to certain cognitive and communicative skills, which from a medical model perspective, can be impaired or affected in DLD and autism. This perspective therefore may explain why DLD and autism have been the most frequently explored conditions in relation to code-switching. As reported above, the increased presence of studies on autism is also linked to the first author's familiarity with the autism and bilingualism literature. Nevertheless, even if we excluded the studies on autism and code-switching which were identified through other methods rather than through database search, autism would still represent one of the most commonly featured conditions in this review. A recent review on multilingualism in neurodevelopmental conditions similarly demonstrated that autism is often explored in relation to exposure to multiple languages (Uljarević et al. 2016). Moreover, Prévost & Tuller (2021) identified a steady interest in the topic over the past few years. Therefore, while our own familiarity with the literature on autism and bilingualism affected the breakdown of identified papers in

8 *Code-switching by individuals with neurodevelopmental conditions*

Table 2, it is safe to say that autism is often more frequently studied in multilingual populations in comparison to many other neurodevelopmental conditions. With these findings in mind, there is a need for exploration of code-switching practices in conditions beyond language impairment and autism.

Despite identifying 31 studies on code-switching in neurodevelopmental conditions, only 58% of them had code-switching as the main focus. Of 13 studies which explored code-switching sporadically, only 1 was identified via the database search. This illustrates the importance of going beyond the database search when conducting scoping reviews, as some of the target review research aims can be addressed only peripherally in certain studies and therefore easily missed out via the database search.

5.2 Demographic (under)representation

There is a striking similarity between the investigation of code-switching in neurodevelopmental conditions and general explorations of bilingualism in neurodivergent individuals. Specifically, the topic remains explored primarily in Western countries and in Indo European languages. In this review, most of the participants resided in North America and in Western Europe, and all but 4 studies included (at least some) bilinguals speaking English in combination with other languages. This finding is partially biased due to the authors' inclusion of literature published in languages that they could understand (English, Galician, Italian, Serbo-Croatian, Spanish). Nevertheless, this finding is particularly concerning as most regions of the world in which multilingualism happens to be frequent and/or code-switching happens to be the norm of communication are not represented in the research with neurodivergent individuals published so far.⁹

While for the purposes of summarising the data related to sex and gender we classified all participants under the labels male and female, we find that across the studies there is no unanimous way of reporting on gender and sex, which leads to lack of clarity in describing the demographic information. Specifically, 16% of studies did not provide information on gender or sex, while only about 39% of studies referred to gender or sex explicitly. Importantly, 45% of studies did not explicitly mention whether data was collected on gender or on sex. Some similar findings have been identified across other reviews related to neurodivergent individuals. For instance, [Girolamo, Shen, et al. \(2023\)](#) found that among the

⁹Examples of code-switching practices of such communities have been described in the work by authors such as [Balam & de Prada Pérez \(2016\)](#) in Belize, [Aboh \(2020\)](#) in Sub-Saharan Africa, [Lipski \(2017\)](#) in Ecuador, among many others.

studies on language impairment in autistic school-aged children which reported information on gender/sex, 41% did not specify whether they reported gender or sex data. This calls for more clarity in reporting practices in order to represent the identity of participants more accurately, as well as allow for precise summaries in studies that review the literature on neurodivergent populations.

Looking at the age of participants whose code-switching practices have been explored, apart from noting that the majority of studies (74%) have been conducted with children (i.e., in individuals under 18 years of age), a striking finding is that the studies related to children included only younger participants. Specifically, all studies focusing only on children had participants with the age of 12 or under. This mirrors the findings by [Prévost & Tuller \(2021\)](#) who identified that explorations of bilingualism and autism are severely under-represented among teenagers. Our findings extend this point to other neurodevelopmental conditions as well as to adult neurodivergent bilinguals, who are also under-represented in our corpus of studies. These results are likely driven by the assumptions that code-switching might cause confusion in neurodivergent children as they develop - consequently, the focus of most studies remains on younger children and their language development. However, considering that code-switching is often a community norm for many multilinguals across their lifespan (e.g., see [Parafita Couto & Balam forthcoming](#)), more work is required across age groups to understand the dynamic nature of neurodivergent bilingualism both from the personal and the linguistic perspective - for instance, what role does code-switching play across lifespan of neurodivergent bilinguals, how does it link to their identity, and do linguistic changes in code-switching patterns emerge at different ages or across various generations of speakers (for some work on this topic in neurotypical bilinguals, see for instance [Georgalidou et al. \(2010\)](#)). Importantly, grasping the nature of code-switching practices among older children and in adults can also inform practitioners more accurately in terms of life outcomes of neurodivergent bilinguals related to their linguistic identity and characteristics.

The final variable which we address in terms of representation across the reviewed studies includes SES. We acknowledge the limitation of our approach in classifying studies on a scale from low to high SES based on certain estimates used across research. For instance, to enable the summary of the data we made the decisions such as classifying studies where the caregivers have predominantly any higher education degree as high SES. However, while holding a university degree does not necessarily imply high SES, in multilingual contexts linked to migration the connection between a university degree and high SES is even weaker. Recent data from the UK context show that 20% of UK born

8 Code-switching by individuals with neurodevelopmental conditions

workers are overqualified for what is labelled as low or medium-low skilled jobs (Fernández-Reino & Brindle 2024). This percentage increases to 26% for all foreign-born workers, and to 27%-29% for workers from Central and South America, European Union (EU) as a whole, Middle East and North Africa and Central Asia, and Sub-Saharan Africa. The increase goes to 39%-44% when it comes to workers from EU-2 countries, Pakistan and other South Asian countries, and EU-8 countries (Fernández-Reino & Brindle 2024).¹⁰ Among the studies in this review, 13 studies (42%) included education as an indicator of SES, 7 of which in combination with other estimates. Considering the data from Fernández-Reino & Brindle (2024), research on multilingualism should avoid relying solely on education as an SES estimate. Implementing the approach by Girolamo, Shen, et al. (2023), who ascribed higher quality to studies which reported SES using more than one indicator, we find that only 8/31 studies (26%) did so in the present review (Angulo Jiménez 2020, Cohen et al. 2025, Gross & Castilla-Earls 2023, Gutiérrez-Clellen et al. 2009, Kapantzoglou et al. 2021, Sheng et al. 2012, Yu 2013). While it remains challenging to outline recommendations on which SES estimates are more adequate for specific contexts, moving beyond a single variable as an SES estimate needs to be prioritised in future research for a more accurate representation of investigated populations.

5.3 Characteristics of neurodivergent code-switching

We now turn to discussing one of the main aims of this scoping review - is code-switching by neurodivergent individuals different to that of neurotypical individuals, and if so, what implications does this have? We interpret these findings in relation to methodological characteristics of reviewed papers, as well as the identified attitudes and professional advice on code-switching in neurodivergent contexts. Of the 18 studies which included some sort of comparison in code-switching between neurodivergent and neurotypical individuals, 9 (50%) identified no or predominantly no differences, 5 (28%) identified (predominantly) different code-switching patterns, 3 (17%) included both similarities and differences, and 1 (6%) produced inconclusive results. As noted in the results section, identified differences in code-switching patterns do not necessarily imply a deficit. Importantly, there is a need to contextualise these findings in terms of existing stigma towards code-switching, especially in a neurodivergent context. In particular, most studies to date identified a negative attitude towards code-switching and negative advice received from professionals, such as educators or speech

¹⁰EU-2 countries include Romania and Bulgaria, while EU-8 countries include Poland, Lithuania, Czech Republic, Hungary, Slovakia, Slovenia, Estonia, and Latvia.

and language therapists. Considering that such stigmatisation of code-switching impacts communicative practices of multilingual families, it likely affects how neurodivergent individuals code-switch. Specifically, preventing exposure to the code-switching community norms (in terms of frequency, types of code-switches, etc.) could lead to different patterns in neurodivergent code-switching. Consequently, any differences in code-switching identified across studies in this review could be related to changes in family communicative practices based on code-switching stigmatisation or pathologization. This further reinforces a need for explorations of neurodivergent code-switching in communities in which multilingualism is the norm and in which stigma about code-switching is non-existent or lower than in the West European and North American contexts, which were overrepresented in this review.

We note that our conservative approach to data extraction on attitudes and advice on code-switching likely impacted the representation of negative views and recommendations. Specifically, we counted instances of attitudes and professional advice only when language mixing was explicitly mentioned. However, negative advice regarding bilingualism in general (e.g., speaking to the child only in the societal language and dropping the home languages) is commonly identified in studies on neurodivergent bilinguals (Uljarević et al. 2016). Considering this, in the present review we focused only on advice and attitudes towards language mixing explicitly. Nevertheless, we acknowledge that professional advice to use only the societal language with the child inevitably implies avoiding code-switching. This suggests that negative attitudes/advice on code-switching are likely more prevalent than what we report. Consequently, any observed differences in neurodivergent code-switching need to be interpreted cautiously. Our conclusion based on the studies in this review is that observed differences in neurodivergent code-switching do not imply deficit or a reason for concern.

How generalisable these findings are also depends on methodological characteristics of the observed studies. Taking into account the ways in which code-switching has been explored across the studies, only 19% of studies used spontaneous speech data and additional 13% of studies used a combination of spontaneous speech and a semi-experimental task. While semi-experimental data can be reflective of code-switching community norms, spontaneous speech data represents the most accurate description of community practices and it facilitates the understanding of data obtained through (semi)experimental approaches (e.g., van Osch et al. 2023, Parafita Couto et al. 2024, Vaughan-Evans et al. 2020). In the present review, 19% of studies used a purely experimental approach in exploring code-switching in neurodivergent bilinguals. Considering the impact that this research can have on the decisions that multilingual families make about language

8 *Code-switching by individuals with neurodevelopmental conditions*

practices with their neurodivergent family members, we invite researchers to be cautious in their study designs and interpretations of their findings. For instance, using sentence repetition tasks or repeated word association tasks (such as in Soesman et al. (2022) or in Sheng et al. (2012)), cued picture selection or naming tasks (as in King et al. (2021) or Snijders et al. (2025)), acceptability judgement and sentence completion tasks (such as in Liceras & Garcia-Alcaraz (2019)) does not necessarily reflect ecologically valid code-switching practices. In fact, some of these tasks (e.g., cued picture naming task in Snijders et al. (2025)) are better described as verbal cognitive rather than linguistic tasks. Therefore, identifying differences between neurodivergent and neurotypical bilinguals by using solely experimental paradigms could lead to increased concerns about language mixing practices of neurodivergent bilinguals, while in fact such tasks do not necessarily reflect the true nature of their bilingual competencies or the role that code-switching can play in their communicative practices or in creating bonds with their families or communities.

A part of the problem is related to the terminology used when exploring code-switching. Due to the lack of consensus on code-switching terminology in the field, and for the purposes of scoping any relevant work on this topic, in this review we used the terms such as code-switching, language mixing, code mixing (and others) interchangeably. While we defined code-switching in the introductory section, it is clear that the work reviewed in this paper goes beyond that phenomenon and in some cases (e.g., Snijders et al. 2025), it includes work on cognitive skills rather than what we consider code-switching. While the debates on differences and similarities between the used terms and phenomena will inevitably carry on and change over time, we invite researchers to continue (and in cases when they have not done so, start) describing in detail the phenomenon under investigation to enable accurate comparability in research. In addition, we note that the lack of consensus on code-switching definitions seems to exist in general (e.g., Gullberg et al. 2009) and not only when exploring code-switching in individuals with neurodevelopmental conditions. While the comparison of code-switching definitions across the studies was beyond the scope of this review (see the aims outlined in the pre-registered protocol: Kaščelan & Parafita Couto (2022)), we encourage researchers to pursue this research question in future reviews on the topic. Importantly, we also invite researchers to make an additional effort to communicate or clarify their work to relevant stakeholders (e.g., multilingual families, neurodivergent individuals, speech and language therapists, teachers, medical experts) in such a way that prevents findings on related but different phenomena from being misused in advising multilingual families

of neurodivergent individuals to drop their home languages or ban/restrict code-switching practices in their interactions.

5.4 Generalisability of findings

We now turn to a common trend of small samples in research on neurodivergent bilinguals and the implications for finding generalisability. In addition to the fact that code-switching in neurodivergent bilinguals is under-explored, we found that only six studies included samples of more than 20 participants. This corroborates a recent review on bilingualism and autism research, which identified small sample sizes across quantitative studies (Prévost & Tuller 2021). In light of these circumstances, there is a question of how generalisable quantitative studies with smaller samples are. As suggested by Girolamo, Shen, et al. (2023), working with small and heterogeneous groups (in this case heterogeneous both in terms of neurodivergent and bi/multilingual status, as well as potentially racial, gender/sex, or socioeconomic characteristics) requires selecting analytic approaches more adequate for such samples (e.g., Bayesian approaches). We also note that the application of qualitative methods could further illuminate our understanding of neurodivergent bi/multilingual's lived experiences in ways not necessarily captured with purely quantitative perspectives (e.g., see Howard et al. (2019) on using Interpretative Phenomenological Analysis in autism research). The combination of quantitative and qualitative approaches is particularly important to consider when exploring community norms of code-switching practices. Specifically, as variation in code-switching community norms can be expected (see Deuchar 2020, Parafita Couto & Balam forthcoming), relying on benefits of mixed methods approaches can help us understand how the heterogeneous nature of neurodivergent bilingualism interacts with established community norms.

The issue of smaller samples and lack of generalisability could somewhat be mitigated following suggestions by Prévost & Tuller (2021) on large-scale inter-lab collaborations. In code-switching research with neurodivergent bilinguals, this could include designing and sharing a protocol for spontaneous speech studies, which would be used across various communities/locations. Apart from a general agreement to collect spontaneous speech data, the protocol could include a list of tools to use for obtaining relevant background information about participants' conditions and their bi/multilingual experiences. Taking into account the diversity and lack of equivalency between the questionnaires used to document bi/multilingualism (see Kaščelan et al. (2022) for questionnaires used with children and Dass et al. (2024) for questionnaires used with adults), it would be important to use a set of shared tools across labs to ensure data comparability.

8 *Code-switching by individuals with neurodevelopmental conditions*

Projects such as Many Babies¹¹ represent a useful example of how such large-scale inter-lab collaborations could be established and navigated.

Inter-lab collaborations would likely contribute to more generalisability through the creation of aggregated datasets, provided the full data availability. However, access to existing corpora of spontaneous speech of bi/multilinguals is a commonly documented issue in code-switching research even in neurotypical samples (Deuchar 2020, Parafita Couto et al. 2023, forthcoming). We strongly agree with the call for making (anonymised) corpora of spontaneous speech fully available. However, care must be taken when making the data accessible to the public particularly when working with vulnerable groups (e.g., neurodivergent individuals, minoritised speakers). Neurotypical researchers are automatically outsiders to neurodivergent communities that they work with. This makes it even more important to approach such projects in a participatory way through the creation of trust and community partnerships. As outlined in Girolamo, Ghali, et al. (2023), a part of this process includes dynamic and interactive informed consent and assent, which as an iterative process would give participants a higher level of control of their own data, while ultimately also making the datasets available to the public whenever possible. Taking such a community-based approach takes years to develop (e.g., see Girolamo, Ghali, et al. 2023 for an example of such a project including a 7-year timeline); nevertheless, a community-based approach enables removal of barriers in recruitment of neurodivergent bilinguals, it helps co-construct accurate narratives about their participation and languaging, and it contributes to responsible data sharing as well as to research transparency. This approach reinforces the importance of slow science for responsible and high-quality research (Fletcher-Watson et al. 2019, Girolamo, Ghali, et al. 2023) and has implications for funding bodies which often fund only short-term projects. Importantly, slow science requires opposing policies of the neoliberal university and the constant requirement for researchers to compete for rankings and publish at a fast pace. Without a culture in which researchers are given time and resources required to conduct studies that include community-based approaches, the work with neurodivergent bilinguals will likely continue facing similar limitations as it has had in the up-to-date research presented in this review and elsewhere (see Prévost & Tuller 2021).

5.5 Implications for practitioners and multilingual families

We now turn to the implications for practitioners (primarily speech and language therapists and educators working with neurodivergent individuals). Some

¹¹See <https://manybabies.org/index>

of these points are also relevant to neurodivergent individuals and their families. Overall, the review indicates that code-switching does not pose a difficulty or threat for a neurodivergent individual's development or language use. On the contrary, across several studies, neurodivergent individuals have used code-switching as a demonstration of their communicative competences (e.g., [Angulo Jiménez 2020](#), [Ponce-Lawler 2017](#), [Yu 2016](#)). Importantly, any observed differences in code-switching patterns between neurodivergent and neurotypical individuals need to be interpreted in light of contextual factors (i.e., predominantly negative attitudes and advice on code-switching) as these might affect the code-switching characteristics of neurodivergent individuals (e.g., the code-switching frequency, the types of code-switches, the contexts in which they code-switch). Additionally, individual variations in code-switching patterns can be expected among neurotypical individuals as well (e.g., [Cohen et al. 2025](#)). Therefore, when variation in code-switching emerges among neurodivergent individuals, it should not be pathologised by default. Considering this, we advise practitioners not to discourage multilingual families from code-switching with their neurodivergent family members, especially when code-switching is a part of their family's/community's communicative practice. These recommendations are the extension to the emerging evidence in bilingualism literature, which suggests that being neurodivergent or having a neurodivergent family member should not imply opting for monolingualism (e.g., [Digard et al. 2023](#), [Prévost & Tuller 2021](#), [Uljarević et al. 2016](#)). This is particularly important considering the evidence about the link between maintaining family multilingualism and family wellbeing ([Müller et al. 2020](#)).

Nevertheless, we also point practitioners to cases when neurodivergent individuals express negative attitudes to code-switching or to the use / exposure to one of their languages. For instance, a case study of a child with stutter identified the child's negative attitude towards using the home language ([Karniol 1992](#)). As a part of evidence-based practice, it is critical that speech and language therapists consider the following three elements: their clinical expertise, the service user's preferences and values, and scientific evidence ([Health and Care Professions Council 2023](#), [Royal College of Speech and Language Therapists n.d.](#)). Therefore, when individuals show reluctance to use all of their languages (which likely includes avoiding code-switching), the healthcare professionals should not aim to enforce communicative practices not preferred by the service users. Instead, they should present the available evidence on multilingualism (which could address possible stigma associated with multilingualism and code-switching) so that neurodivergent individuals and their families can make informed decisions about their preferences.

8 *Code-switching by individuals with neurodevelopmental conditions*

The research on stigma towards code-switching and multilingualism in general requires further investigation. We suspect that identifying predominantly negative advice and attitudes on code-switching in this review is linked to the fact that most of the reviewed studies were conducted in countries with monolingualism-centric views. Apart from continuing the investigation of this topic in the communities included in this review, it is imperative to learn about the code-switching phenomenon (especially in relation to neurodivergent individuals) from underexplored parts of the world where multilingualism is the norm. We also invite researchers exploring this topic to make a clear distinction between attitudes/advice on multilingualism and attitudes/advice on code-switching. Specifically, while in some monolingualism-centric countries, attitudes on multilingualism might be improving (e.g., see [Digard et al. 2023](#)), the stigma on code-switching could still remain due to the pressure imposed by standard language ideologies and their prescriptivist, puristic and racialised views of communicative practices.

6 Concluding remarks

This scoping review outlines the existing findings on code-switching in bi/multilinguals with neurodevelopmental conditions. Generally, we do not identify evidence of harmful effects of code-switching for neurodivergent individuals. In light of this, we make recommendations for practitioners, especially speech and language therapists and educators working with neurodivergent individuals. We do however find that this topic remains severely under-explored and the existing evidence suffers from bias in focus (primarily on DLD and autism) and demographic under-representation (i.e., focus has been mostly on younger children, on specific geographic locations and language combinations). We further observe that the lack of terminological consensus on code-switching translates from research with neurotypical individuals onto the research with neurodivergent population. We invite researchers to consider how they communicate their findings (especially in relation to neurodivergent population) as some of the used experimental approaches explored related constructs to code-switching but not code-switching per se. Therefore, these findings should not be the core literature informing recommendations on code-switching practices in neurodivergent contexts. In light of these downsides, we discuss generalisability of existing findings and offer suggestions about methodological and practical issues in exploring this topic.

We would also like to highlight that while our key search terms included terminology related to code-blending for the purposes of including work from sign

Draško Kaščelan & María del Carmen Parafita Couto

languages, we did not succeed in identifying any relevant work. Consequently, this review is focused on the spoken modality only, so we encourage researchers to pursue this topic taking into account multimodal perspectives. Finally, this review focused on neurodevelopmental conditions. Considering that bilingualism is often a lifelong and a dynamic experience, there is a need to explore code-switching practices in acquired conditions as well. Relevant discussions of code-switching in conditions such as aphasia can be found in [Aboh \(2020\)](#).

Supplementary materials

The datafile can be accessed on the project OSF page: <https://osf.io/4nwgdl/>

Note: The studies included in the scoping review are marked with an asterisk.

References

- Aboh, Enoch O. 2020. Lessons from neuro-(a)-typical brains: Universal multilingualism, code-mixing, recombination, and executive functions. *Frontiers in Psychology* 11. 488. DOI: [10.3389/fpsyg.2020.00488](https://doi.org/10.3389/fpsyg.2020.00488).
- Aguilar-Mediavilla, Eva, Alberto Sánchez-Pedroche, Lucía Buil-Legaz, Josep A. Pérez-Castelló & Daniel Adrover-Roig. 2024. Code-switching and code-mixing in bilingual Spanish–Catalan children with and without developmental language disorder. In Vicenç Torrens (ed.), *Language acquisition in Romance languages* (Bilingual Processing and Acquisition 18), 283–305. Amsterdam: John Benjamins. DOI: [10.1075/bpa.18.12san](https://doi.org/10.1075/bpa.18.12san).
- American Psychiatric Association. 2013. *DSM-5. Diagnostic and statistical manual of mental disorders*. 5th edn. Washington, D.C.: American Psychiatric Association. DOI: [10.1176/appi.books.9780890425596](https://doi.org/10.1176/appi.books.9780890425596).
- Angulo Jiménez, Henry Giovanny. 2020. “I was like ‘oh my God, los dos’”: Code-switching in conversations with a bilingual adult on the autism spectrum. University of Illinois at Urbana-Champaign. (Doctoral dissertation). <https://hdl.handle.net/2142/107884>.
- Angulo-Jiménez, Henry & Laura DeThorne. 2024. Discourse-related code-switching in conversations with a bilingual autistic adult. *Journal of Interactional Research in Communication Disorders* 15(2). 119–146. DOI: [10.1558/jircd.25657](https://doi.org/10.1558/jircd.25657).
- Balam, Osmer. 2021. Beyond differences and similarities in codeswitching and translanguaging research. *Belgian Journal of Linguistics* 35(1). 76–103. DOI: [10.1075/bjl.00065.bal](https://doi.org/10.1075/bjl.00065.bal).

8 Code-switching by individuals with neurodevelopmental conditions

- Balam, Osmer & Ana de Prada Pérez. 2016. On the productive use of ‘hacer + V’ in Northern Belize bilingual/trilingual codeswitching. In Rosa E. Guzzardo Tamargo, Catherine M. Mazak & Maria del Carmen Parafita Couto (eds.), *Spanish-English codeswitching in the Caribbean and the US* (Issues in Hispanic and Lusophone Linguistics 11), 261–279. Amsterdam: John Benjamins. DOI: [10.1075/ihll.11.11bal](https://doi.org/10.1075/ihll.11.11bal).
- Blom, Elma, Gülşah Yazıcı, Tessel Boerma & Merel van Witteloostuijn. 2024. A longitudinal study of Turkish-Dutch children’s language mixing in single-language settings: Language status, language proficiency, cognitive control and developmental language disorder. *Cognitive Development* 71. 101481. DOI: [10.1016/j.cogdev.2024.101481](https://doi.org/10.1016/j.cogdev.2024.101481).
- Cohen, Shana R., A. Wishard Guerra, Jessica Miguel, Kristen Bottema-Beutel & Gabrielle Oliveira. 2025. *Hablando* at home: Examining the interactional resources of a bilingual autistic child. *Journal of Child Language* 52(1). 135–157. DOI: [10.1017/S0305000923000600](https://doi.org/10.1017/S0305000923000600).
- Dass, Ronessa, Irina Smirnova-Godoy, Olivia McColl, John G. Grundy, Gigi Luk & John A. E. Anderson. 2024. A content overlap analysis of bilingualism questionnaires: Considering diversity. *Bilingualism: Language and Cognition* 27(4). 744–750. DOI: [10.1017/S1366728923000767](https://doi.org/10.1017/S1366728923000767).
- Deuchar, Margaret. 2020. Code-switching in linguistics: A position paper. *Languages* 5(2). 22. DOI: [10.3390/languages5020022](https://doi.org/10.3390/languages5020022).
- Deuchar, Margaret, Peredur Davies, Jon Russell Herring, Maria del Carmen Parafita Couto & Diana Carter. 2014. Building bilingual corpora. In Enlli Môn Thomas & Ineke Mennen (eds.), *Advances in the study of bilingualism*, 93–110. Bristol: Multilingual Matters. DOI: [10.21832/9781783091713-008](https://doi.org/10.21832/9781783091713-008).
- Digard, Bérangère Galadriel, Ellie Johnson, Draško Kaščelan & Rachael Davis. 2023. Raising bilingual autistic children in the UK: At the intersection between neurological and language diversity. *Frontiers in Psychiatry* 14. DOI: [10.3389/fpsyt.2023.1250199](https://doi.org/10.3389/fpsyt.2023.1250199).
- Fernandez y Garcia, Erik, Joshua Breslau, Robin Hansen & Elizabeth Miller. 2012. Unintended consequences: An ethnographic narrative case series exploring language recommendations for bilingual families of children with autistic spectrum disorders. *Journal of Medical Speech-Language Pathology* 20. 10–16.
- Fernández-Reino, Mariña & Ben Brindle. 2024. *Migrants in the UK labour market: An overview*. Migration Observatory briefing. Oxford: COMPAS, University of Oxford. <https://migrationobservatory.ox.ac.uk/resources/briefings/migrants-in-the-uk-labour-market-an-overview/>.

Draško Kaščelan & María del Carmen Parafita Couto

- Fletcher-Watson, Sue, Jon Adams, Kabie Brook, Tony Charman, Laura Crane, James Cusack, Susan Leekam, Damian Milton, Jeremy R Parr & Elizabeth Pellicano. 2019. Making the future together: Shaping autism research through meaningful participation. *Autism* 23(4). 943–953. DOI: [10.1177 / 1362361318786721](https://doi.org/10.1177/1362361318786721).
- Georgalidou, Marianthi, Hasan Kaili & Aytac Celtek. 2010. Code alternation patterns in bilingual family conversation: A conversation analysis approach. *Journal of Greek Linguistics* 10(2). 317–344. DOI: [10.1163/156658410X531401](https://doi.org/10.1163/156658410X531401).
- Girolamo, Teresa, Samantha Ghali & Inge-Marie Eigsti. 2023. A community-based approach to longitudinal language research with racially and ethnically minoritized autistic young adults: Lessons learned and new directions. *American Journal of Speech-Language Pathology* 32(3). 977–988. DOI: [10.1044/2023_AJSLP-22-00341](https://doi.org/10.1044/2023_AJSLP-22-00341).
- Girolamo, Teresa, Lue Shen, Amalia Monroe Gulick, Mabel L. Rice & Inge-Marie Eigsti. 2023. Studies pertaining to language impairment in school-age autistic individuals underreport participant sociodemographics: A systematic review. *Autism* 27(8). 2218–2240. DOI: [10.1177/13623613231166749](https://doi.org/10.1177/13623613231166749).
- Gross, Megan C. & Anny Castilla-Earls. 2023. Code-switching during narratives by bilingual children with and without developmental language disorder. *Language, Speech, and Hearing Services in Schools* 54(3). 996–1019. DOI: [10.1044 / 2023_LSHSS-22-00149](https://doi.org/10.1044/2023_LSHSS-22-00149).
- Gross, Megan C. & Margarita Kaushanskaya. 2022. Language control and code-switching in bilingual children with developmental language disorder. *Journal of Speech, Language, and Hearing Research* 65(3). 1104–1127. DOI: [10.1044/2021_JSLHR-21-00332](https://doi.org/10.1044/2021_JSLHR-21-00332).
- Gullberg, Marianne, Peter Indefrey & Pieter Muysken. 2009. Research techniques for the study of code-switching. In Almeida Jacqueline Toribio & Barbara E. Bullock (eds.), *The Cambridge handbook of linguistic code-switching* (Cambridge Handbooks in Language and Linguistics), 21–39. Cambridge: Cambridge University Press. DOI: [10.1017/CBO9780511576331.003](https://doi.org/10.1017/CBO9780511576331.003).
- Gutiérrez-Clellen, Vera F., Gabriela Simon-Cereijido & Angela Erickson Leone. 2009. Code-switching in bilingual children with specific language impairment. *International Journal of Bilingualism* 13(1). 91–109. DOI: [10.1177 / 1367006909103530](https://doi.org/10.1177/1367006909103530).
- Health and Care Professions Council. 2023. *The standards of proficiency for speech and language therapists*. <https://www.hcpc-uk.org/standards/standards-of-proficiency/speech-and-language-therapists/>.

8 Code-switching by individuals with neurodevelopmental conditions

- Howard, Katie, Napoleon Katsos & Jenny Gibson. 2019. Using interpretative phenomenological analysis in autism research. *Autism* 23(7). 1871–1876. DOI: [10.1177/1362361318823902](https://doi.org/10.1177/1362361318823902).
- Howard, Katie B., Napoleon Katsos & Jenny L. Gibson. 2021. Practitioners' perspectives and experiences of supporting bilingual pupils on the autism spectrum in two linguistically different educational settings. *British Educational Research Journal* 47(2). 427–449. DOI: [10.1002/berj.3662](https://doi.org/10.1002/berj.3662).
- Iluz-Cohen, Peri & Joel Walters. 2012. Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. *Bilingualism: Language and Cognition* 15(1). 58–74. DOI: [10.1017/S1366728911000538](https://doi.org/10.1017/S1366728911000538).
- Jegatheesan, Brinda. 2011. Multilingual development in children with autism: Perspectives of South Asian Muslim immigrant parents on raising a child with a communicative disorder in multilingual contexts. *Bilingual Research Journal* 34(2). 185–200. DOI: [10.1080/15235882.2011.597824](https://doi.org/10.1080/15235882.2011.597824).
- Kapantzoglou, Maria, Julie Esparza Brown, Lauren M. Cycyk & Gerasimos Fergadiotis. 2021. Code-switching and language proficiency in bilingual children with and without developmental language disorder. *Journal of Speech, Language, and Hearing Research* 64(5). 1605–1620. DOI: [10.1044/2020_JSLHR-20-00182](https://doi.org/10.1044/2020_JSLHR-20-00182).
- Karniol, Rachel. 1992. Stuttering out of bilingualism. *First Language* 12(36). 255–283. DOI: [10.1177/014272379201203604](https://doi.org/10.1177/014272379201203604).
- Kaščelan, Draško, Philippe Prévost, Ludovica Serratrice, Laurie Tuller, Sharon Unsworth & Cécile De Cat. 2022. A review of questionnaires quantifying bilingual experience in children: Do they document the same constructs? *Bilingualism: Language and Cognition* 25(1). 29–41. DOI: [10.1017/S1366728921000390](https://doi.org/10.1017/S1366728921000390).
- Kaščelan, Draško & Maria del Carmen Parafita Couto. 2022. *Code-switching by individuals with neurodevelopmental conditions: A scoping review protocol*. DOI: [10.17605/OSF.IO/JPW3X](https://doi.org/10.17605/OSF.IO/JPW3X).
- Kay-Raining Bird, Elizabeth, Erin Lamond & Jeanette Holden. 2012. Survey of bilingualism in autism spectrum disorders. *International Journal of Language & Communication Disorders* 47(1). 52–64. DOI: [10.1111/j.1460-6984.2011.00071.x](https://doi.org/10.1111/j.1460-6984.2011.00071.x).
- King, Marika R., MaryAnn Ronski & Rose A. Sevcik. 2021. Language differentiation using augmentative and alternative communication: An investigation of Spanish–English bilingual children with and without language impairments. *American Journal of Speech-Language Pathology* 30(1). 89–104. DOI: [10.1044/2020_AJSLP-20-00030](https://doi.org/10.1044/2020_AJSLP-20-00030).
- Kremer-Sadlik, Tamar. 2005. To be or not to be bilingual: Autistic children from multilingual families. In James Cohen, Kara T. McAlister, Kellie Rolstad & Jeff

- MacSwan (eds.), *Proceedings of the 4th International Symposium on Bilingualism*, 1225–1234. Somerville: Cascadilla Press.
- Li, Wei. 2018. Translanguaging as a practical theory of language. *Applied Linguistics* 39(1). 9–30. DOI: [10.1093/applin/amx039](https://doi.org/10.1093/applin/amx039).
- Liceras, Juana M. & Estela Garcia-Alcaraz. 2019. Grammatical gender in atypical language development. *Journal of Monolingual and Bilingual Speech* 1(2). 225–247. DOI: [10.1558/jmbs.v1i2.11878](https://doi.org/10.1558/jmbs.v1i2.11878).
- Lipski, John M. 2017. Language switching constraints: More than syntax? Data from Media Lengua. *Bilingualism: Language and Cognition* 20(4). 722–746. DOI: [10.1017/S1366728916000468](https://doi.org/10.1017/S1366728916000468).
- Mammolito, Angela P. 2015. *Code-switching and grammatical error rates of Spanish-English bilingual school-aged children with a language disorder*. Tallahassee: Florida State University. (Doctoral dissertation). <https://www.proquest.com/openview/b4c3f3df821c252b4e250a62ceca5c60/>.
- Müller, Lisa-Maria, Katie Howard, Elspeth Wilson, Jenny Gibson & Napoleon Katsos. 2020. Bilingualism in the family and child well-being: A scoping review. *International Journal of Bilingualism* 24(5–6). 1049–1070. DOI: [10.1177/1367006920920939](https://doi.org/10.1177/1367006920920939).
- Munn, Zachary, Danielle Pollock, Hanan Khalil, Lyndsay Alexander, Patricia McInerney, Christina M. Godfrey, Micah Peters & Andrea C. Tricco. 2022. What are scoping reviews? Providing a formal definition of scoping reviews as a type of evidence synthesis. *JBIE Evidence Synthesis* 20(4). 950. DOI: [10.11124/JBIES-21-00483](https://doi.org/10.11124/JBIES-21-00483).
- Myers-Scotton, Carol. 1983. The negotiation of identities in conversation: A theory of markedness and code choice. *International Journal of the Sociology of Language* 1983(44). 115–136. DOI: [10.1515/ijsl.1983.44.115](https://doi.org/10.1515/ijsl.1983.44.115).
- O'Toole, Ciara & Tina M Hickey. 2013. Diagnosing language impairment in bilinguals: Professional experience and perception. *Child Language Teaching and Therapy* 29(1). 91–109. DOI: [10.1177/0265659012459859](https://doi.org/10.1177/0265659012459859).
- Oudet, Sarah, Katie Howard & Stephanie Durrleman. 2022. Early years autism and bilingualism: An interpretative phenomenological analysis of parent perceptions during lockdown. *Autism & Developmental Language Impairments* 7. DOI: [10.1177/23969415221138704](https://doi.org/10.1177/23969415221138704).
- Parafita Couto, Maria del Carmen, Charlotte Pouw, Rodi Laanen & Luis López. 2024. The role of INFL in code-switching: A study of a Papiamentto heritage community in the Netherlands. *Frontiers in Language Sciences* 2. DOI: [10.3389/flang.2023.1288198](https://doi.org/10.3389/flang.2023.1288198).

8 Code-switching by individuals with neurodevelopmental conditions

- Parafita Couto, Maria del Carmen & Osmer Balam. Forthcoming. Types of code-switching and code-switching communities. In Leonardo Cerno, Hans-Jörg Döhla, Miguel Gutiérrez Maté, Robert Hesselbach & Joachim Steffen (eds.), *Contact varieties of Spanish and Spanish-lexified contact varieties: An international handbook* (Handbooks for Linguistics and Communication Science / Handbücher zur Sprach- und Kommunikationswissenschaft). Berlin: Mouton De Gruyter.
- Parafita Couto, Maria del Carmen, Kate Bellamy & Felix K. Ameka. 2023. Theoretical linguistic approaches to multilingual code-switching. In Jennifer Cabrelli, Adel Chaouch-Orozco, Jorge González Alonso, Sergio Miguel Pereira Soares, Eloi Puig-Mayenco & Rothman Jason (eds.), *The Cambridge handbook of third language acquisition* (Cambridge Handbooks in Language and Linguistics), 403–436. Cambridge: Cambridge University Press. DOI: [10.1017/9781108957823.017](https://doi.org/10.1017/9781108957823.017).
- Parafita Couto, Maria del Carmen, Miriam Greidanus Romanelli & Kate Bellamy. Forthcoming. Code-switching at the interface between language, culture, and cognition. *Lapurdum*.
- Pert, Sean. 2007. *Bilingual language development in Pakistani heritage children in Rochdale UK : Intrasentential codeswitching and the implications for identifying specific language impairment*. Newcastle: Newcastle University. (Doctoral dissertation). <http://hdl.handle.net/10443/2230>.
- Peters, Micah D. J., Christina Godfrey, Patricia McNerney, Hanan Khalil, Palle Larsen, Casey Marnie, Danielle Pollock, Andrea C. Tricco & Zachary Munn. 2022. Best practice guidance and reporting items for the development of scoping review protocols. *JBI Evidence Synthesis* 20(4). 953–968. DOI: [10.11124/JBIES-21-00242](https://doi.org/10.11124/JBIES-21-00242).
- Peters, Micah D. J., Casey Marnie, Heather Colquhoun, Chantelle M. Garritty, Susanne Hempel, Tanya Horsley, Etienne V. Langlois, Erin Lillie, Kelly K. O'Brien, Özge Tunçalp, Michael G. Wilson, Wasifa Zarin & Andrea C. Tricco. 2021. Scoping reviews: Reinforcing and advancing the methodology and application. *Systematic Reviews* 10(1). 263. DOI: [10.1186/s13643-021-01821-3](https://doi.org/10.1186/s13643-021-01821-3).
- Peters, Micah D. J., Casey Marnie, Andrea C. Tricco, Danielle Pollock, Zachary Munn, Lyndsay Alexander, Patricia McNerney, Christina M. Godfrey & Hanan Khalil. 2020. Updated methodological guidance for the conduct of scoping reviews. *JBI Evidence Implementation* 19(1). 3. DOI: [10.1097/XEB.000000000000277](https://doi.org/10.1097/XEB.000000000000277).
- Phillips, Shannon & Margaret Deuchar. 2022. The role of the input in the acquisition of code-switching. In Unn Røyneland & Robert Blackwood (eds.), *Mul-*

- lingualism across the lifespan* (Routledge Critical Studies in Multilingualism), 56–79. New York: Routledge. DOI: [10.4324/9781003125815-5](https://doi.org/10.4324/9781003125815-5).
- Pollock, Danielle, Andrea C. Tricco, Micah D. J. Peters, Patricia A. McInerney, Hanan Khalil, Christina M. Godfrey, Lyndsay A. Alexander & Zachary Munn. 2022. Methodological quality, guidance, and tools in scoping reviews: A scoping review protocol. *JBI Evidence Synthesis* 20(4). 1098. DOI: [10.11124/JBIES-20-00570](https://doi.org/10.11124/JBIES-20-00570).
- Ponce-Lawler, Monica Yolanda. 2017. *The use of code-switching in bilingual children with autism spectrum disorder*. El Paso: University of Texas, El Paso. (Doctoral dissertation). <https://scholarworks.utep.edu/dissertations/AAI10281413>.
- Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics* 18(7/8). 581–618. DOI: [10.1515/ling.1980.18.7-8.581](https://doi.org/10.1515/ling.1980.18.7-8.581).
- Prévost, Philippe & Laurice Tuller. 2021. Bilingual language development in autism. *Linguistic Approaches to Bilingualism* 12(1). 1–32. DOI: [10.1075/lab.21018.pre](https://doi.org/10.1075/lab.21018.pre).
- Restrepo, María Adelaida. 2003. Spanish language skills in bilingual children with specific language impairment. In Silvana Montrul & Francisco Ordóñez (eds.), *Linguistic theory and language development in Hispanic languages: Papers from the 5th Hispanic Linguistics Symposium and the 4th Conference on the Acquisition of Spanish and Portuguese*, 365–374. Somerville: Cascadia Press.
- Royal College of Speech and Language Therapists. N.d. *An introduction to evidence-based practice*. <https://www.rcslt.org/members/research/evidence-based-practice/>.
- Sanz-Torrent, Mònica, Iris Badia & Miquel Serra. 2007. Contributions from bilingual Specific Language Impairment in Catalan and Spanish to the understanding of typical and pathological language acquisition. In Carmen Pérez-Vidal, Maria Juan-Garau & Aurora Bel (eds.), *A portrait of the young in the new multilingual Spain*. Bristol: Multilingual Matters.
- Sheng, Li, Elizabeth D. Peña, Lisa M. Bedore & Christine E. Fiestas. 2012. Semantic deficits in Spanish–English bilingual children with language impairment. *Journal of Speech, Language, and Hearing Research* 55(1). 1–15. DOI: [10.1044/1092-4388\(2011/10-0254](https://doi.org/10.1044/1092-4388(2011/10-0254).
- Snijders, Vera, Merel van Witteloostuijn, Tessel Boerma, Mona Timmermeister & Elma Blom. 2025. Effects of dominance on language switching: A longitudinal study of Turkish–Dutch children with and without developmental language disorder. *Bilingualism: Language and Cognition* 28(1). 272–285. DOI: [10.1017/S1366728924000427](https://doi.org/10.1017/S1366728924000427).

8 Code-switching by individuals with neurodevelopmental conditions

- Snijders, Vera Elisabeth, Merel van Witteloostuijn, Ora Oudgenoeg-Paz & Elma Blom. 2023. *Language mixing in young multilingual children and its correlates: A systematic review*. PROSPERO CRD42023484906. <https://www.crd.york.ac.uk/PROSPERO/view/CRD42023484906>.
- Soesman, Aviva, Joel Walters & Sveta Fichman. 2022. Language control and intra-sentential codeswitching among bilingual children with and without developmental language disorder. *Languages* 7(4). 249. DOI: [10.3390 / languages7040249](https://doi.org/10.3390/languages7040249).
- Treffers-Daller, Jeanine. 2019. What defines language dominance in bilinguals? *Annual Review of Linguistics* 5. 375–393. DOI: [10.1146/annurev - linguistics - 011817-045554](https://doi.org/10.1146/annurev-linguistics-011817-045554).
- Tricco, Andrea C., Erin Lillie, Wasifa Zarin, Kelly K. O'Brien, Heather Colquhoun, Danielle Levac, David Moher, Micah D. J. Peters, Tanya Horsley, Laura Weeks, Susanne Hempel, Elie A. Akl, Christine Chang, Jessie McGowan, Lesley Stewart, Lisa Hartling, Adrian Aldcroft, Michael G. Wilson, Chantelle Garritty, Simon Lewin, Christina M. Godfrey, Marilyn T. Macdonald, Etienne V. Langlois, Karla Soares-Weiser, Jo Moriarty, Tammy Clifford, Özge Tunçalp & Sharon E. Straus. 2018. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine* 169(7). 467–473. DOI: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850).
- Truscott, John & Mike Sharwood Smith. 2017. Representation, Processing and Code-switching. *Bilingualism: Language and Cognition* 20(5). 903–916. DOI: [10.1017/S1366728916000742](https://doi.org/10.1017/S1366728916000742).
- Uljarević, Mirko, Napoleon Katsos, Kristelle Hudry & Jenny L. Gibson. 2016. Multilingualism and neurodevelopmental disorders – An overview of recent research and discussion of clinical implications. *Journal of Child Psychology and Psychiatry* 57(11). 1205–1217. DOI: [10.1111/jcpp.12596](https://doi.org/10.1111/jcpp.12596).
- van Osch, Brechje, Maria del Carmen Parafita Couto, Ivo Boers & Bo Sterken. 2023. Adjective position in the code-switched speech of Spanish and Papiamentu heritage speakers in the Netherlands: Individual differences and methodological considerations. *Frontiers in Psychology* 14. DOI: [10.3389/fpsyg.2023.1136023](https://doi.org/10.3389/fpsyg.2023.1136023).
- Vaughan-Evans, Awel, Maria del Carmen Parafita Couto, Bastien Boutonnet, Noriko Hoshino, Peredur Webb-Davies, Margaret Deuchar & Guillaume Thierry. 2020. Switchmate! An electrophysiological attempt to adjudicate between competing accounts of adjective-noun code-switching. *Frontiers in Psychology* 11. DOI: [10.3389/fpsyg.2020.549762](https://doi.org/10.3389/fpsyg.2020.549762).

Draško Kaščelan & María del Carmen Parafita Couto

- Wharton, Robert H., Karen Levine, Elizabeth Miller, Joshua Breslau & Stanley Greenspan. 2000. Children with special needs in bilingual families: A developmental approach to language recommendations. In *ICDL clinical practice guidelines*, 141–151. Bethesda: Interdisciplinary Council on Developmental & Learning Disorders.
- Yim, Odilia & Richard Clément. 2021. Acculturation and attitudes toward code-switching: A bidimensional framework. *International Journal of Bilingualism* 25(5). 1369–1388. DOI: [10.1177/13670069211019466](https://doi.org/10.1177/13670069211019466).
- Yow, Wei Quin, Jessica S. H. Tan & Suzanne Flynn. 2018. Code-switching as a marker of linguistic competence in bilingual children. *Bilingualism: Language and Cognition* 21(5). 1075–1090. DOI: [10.1017/S1366728917000335](https://doi.org/10.1017/S1366728917000335).
- Yu, Betty. 2013. Issues in bilingualism and heritage language maintenance: Perspectives of minority-language mothers of children with autism spectrum disorders. *American Journal of Speech-Language Pathology* 22(1). 10–24. DOI: [10.1044/1058-0360\(2012/10-0078](https://doi.org/10.1044/1058-0360(2012/10-0078)).
- Yu, Betty. 2016. Code-switching as a communicative resource within routine, bilingual family interactions for a child on the autism spectrum. *Perspectives of the ASHA Special Interest Groups* 1(14). 17–28. DOI: [10.1044/persp1.SIG14.17](https://doi.org/10.1044/persp1.SIG14.17).

Current insights into code-switching

This volume presents a series of studies and reviews that provide insights into grammatical variation and the development of code-switching in neurotypical speakers, as well as into language mixing practices among individuals with neurodevelopmental conditions. Across seven chapters, a range of themes is explored, including: community norms and language practices, the influence of caregivers on bilingual development, attitudes and ideologies surrounding bilingualism, code-switching as a communicative resource, and the intersection of neurodiversity and bilingualism. By offering up-to-date perspectives on these topics, the volume builds on existing research into code-switching, such as work by Margaret Deuchar, and it highlights open data initiatives for advancing future research in this area