Lost in Transmission

The role of attrition and input in heritage language development

EDITED BY

Bernhard Brehmer
Jeanine Treffers-Daller

JOHN BENJAMINS

PUBLISHING COMPANY

Lost in Transmission

Studies in Bilingualism (SiBil)

The focus of this series is on psycholinguistic and sociolinguistic aspects of bilingualism. This entails topics such as childhood bilingualism, psychological models of bilingual language users, language contact and bilingualism, maintenance and shift of minority languages, and socio-political aspects of bilingualism.

For an overview of all books published in this series, please see
benjamins.com/catalog/sibil

Volume 59

Lost in Transmission
The role of attrition and input in heritage language development
Edited by Bernhard Brehmer and Jeanine Treffers-Daller

Lost in Transmission

The role of attrition and input
in heritage language development

Edited by

Bernhard Brehmer
University of Konstanz

Jeanine Treffers-Daller
University of Reading

John Benjamins Publishing Company
Amsterdam / Philadelphia

Table of contents

Acknowledgements

Two sides of one coin?

The relevance of first language attrition for the acquisition of heritage languages

Bernhard Brehmer and Jeanine Treffers-Daller
University of Konstanz / University of Reading

The current volume brings together insights from two different, but related fields of bilingualism, namely acquisition and attrition of speakers' first languages in immigration contexts. Although it seems intuitively plausible to link first language attrition, i.e. the "forgetting" of one's first language in an immigrant setting, with the acquisition of minority languages by later generations of immigrants, only very recently have scholars started to empirically investigate how language maintenance and change within the first generation impact on the acquisition of language(s) by representatives of subsequent generations in the new host community.

Language attrition refers to "the non-pathological decrease in a language that had previously been acquired by an individual" (Köpke & Schmid 2004: 5). Thus, first language attrition deals with the temporary or permanent loss of aspects of a native language (L1) by an individual which is triggered mostly due to a change in the linguistic environment or behavior of the speaker. The vast majority of research in L1 attrition concerns late sequential adult bilinguals who left their home country in adulthood (see Schmid & Köpke 2007; Schmid 2011, 2013). Due to the new second language (L2) environment and the need to acquire a new language, the individual experiences a more or less rapid decrease in the relative use of his/her native language. As an eventual consequence of this reduced input and output in the L1, and co-activation of the L1 and the L2, the native language can undergo substantial changes at all linguistic levels (e.g. phonology, morphosyntax, lexicon, semantics, pragmatics, narrative conventions), in reception as well as production (Pavlenko 2004). Whether or not the term attrition should be used to refer only to permanent traces of the L2 in the L1 is controversial. According to Schmid and Köpke (2017) there is a continuum between online/transient and representational/permanent effects of the L2 on the L1, and establishing distinct stages on this continuum is not possible. Therefore, Schmid and Köpke suggest that every bilingual is in fact an L1 attriter.

Even bilinguals who had extensive access to the L1 until adulthood can be seen as "non-native speakers" of his/her native language by monolingual peers. These judgments are "generally based on observed difficulties with lexical retrieval, the use of codeswitching to fill lexical gaps, divergent pronunciation, morphological errors, avoidance of certain structures, and overuse of other structures due to transfer" (Benmamoun, Montrul & Polinsky 2013: 132). However, L1 attrition may already begin in childhood and can be more pronounced if immigration happens before puberty (that is, up to the age of 12). Numerous studies have shown that in the case of child L1 attrition the extent and speed of the loss of structural aspects in the L1 is more severe than in the case of adults who had grown up with only the L1 before emigration (cf. Bylund 2009; Flores 2010, 2012; Montrul 2008 or Pallier 2007).

A gradual shift from one's native language to the majority language of the surrounding community is also a typical feature of heritage speakers. According to one of the many definitions,

> a heritage speaker is an early bilingual who grew up hearing (and speaking) the heritage language (L1) and the majority language (L2) either simultaneously or sequentially in early childhood (that is, roughly up to age 5 […]), but for whom L2 became the primary language at some point during childhood (at, around, or after the onset of schooling). As a result of language shift, by early adulthood a heritage speaker can be strongly dominant in the majority language, while the heritage language will now be the weaker language.
>
> (Benmamoun, Montrul & Polinsky 2013: 133)

Thus, L1 attrition is one of the processes that shape the development of the heritage language and the linguistic profile of heritage speakers over their lifespan. However, in contrast to research on L1 attrition, which is mostly concerned with the native language of bilingual speakers who started learning the language of their new host community only as adults, the focus of heritage language research is on simultaneous and/or early sequential bilinguals who were either already born in the host community or immigrated at a very early age, normally before the onset of schooling (cf., among many others, Kupisch & Rothman 2018; Montrul 2008, 2016; Polinsky 2018; Rothman 2007, 2009; Rothman & Treffers-Daller 2014). Heritage speakers receive reduced input in the heritage language, as this language is generally used within the family only, while the societally dominant language is used for most other domains. Furthermore, heritage speakers often receive no institutional support for the development of their heritage language, as schooling normally takes place exclusively in the majority language. Consequently, the acquisition of literacy skills or structural aspects of the heritage language which are restricted to formal registers of language use depends on measures that the parents undertake to maintain the heritage language in the family or on the possibility to

attend heritage language classes (see Dąbrowska 2013; Kupisch 2013). Given the complex circumstances shaping the process of heritage language acquisition and the multitude of factors involved therein (cf. Montrul 2016; Polinsky 2018 for an overview), there is extensive interindividual variability with regard to proficiency in the heritage language. Polinsky and Kagan (2007) capture this variability by proposing a continuum ranging from basilectal (= low-proficient) heritage speakers who often have only receptive skills in their heritage language to acrolectal (= high-proficient) heritage speakers who are hardly distinguishable from monolingual peers, except for their familiarity with formal registers in the heritage language.

A common trait of both L1 attriters and heritage speakers is that, despite the huge amount of variation between individual heritage speakers, they differ in some respect from monolinguals. In the case of late sequential bilinguals access to L1 lexical items and grammatical structures may be weakened or even lost after several years of reduced exposure to and use of the L1. In those cases where these had been acquired at an earlier stage this process can be considered as language attrition. Furthermore, transfer from structures of the functionally dominant L2 to the L1 is a common by-product of language attrition, leading to convergence between the two languages. However, proving that that the speakers had mastered the structure under focus before the attrition process began is difficult if data are collected at one point in time only. The best way to prove that attrition is at stake is to conduct longitudinal within group studies. Another possible way is to compare child heritage speakers with adult heritage speakers living in the same community. If the property under focus is attested in the younger heritage speakers (e.g. between the ages of 4–6), but not in the older ones (e.g. between the ages of 15–20), attrition is likely to have occurred in the older group. However, given the reduced and variable amount of input in the heritage language, it could also be the case that input was quantitatively not sufficient to establish a stable command of a given property in the heritage language. This is a likely outcome especially with items or properties that are infrequent in the parental input or bound to certain formal registers of the heritage language (e.g. passive constructions or structures related to literacy skills). At best, the speakers catch up with these developmental delays at later ages compared to monolingual controls. Otherwise this leads to what has been called "incomplete" or "arrested" acquisition of the heritage language (cf. Montrul 2008; Polinsky 2006, 2008), resulting in non-native-like attainment in the heritage language with regard to the respective structures.[1] Transfer from the functionally (and often also structurally) dominant majority language might be another factor that accounts for

---

1.   For a substantial criticism on the term "incomplete acquisition" see, among others, Kupisch & Rothman (2018).

the patterns found among heritage speakers. However, the question arises whether these transfers can be directly traced back to processing problems in inhibiting the majority language on the side of the heritage speakers, or whether the transferred structures were already part of the parental input they received. This brings us back to the relationship between L1 attrition and heritage language acquisition.

The crucial point here is that input in the heritage language is usually provided by parents who are themselves (late sequential) bilinguals. Whereas the "incomplete acquisition" and the "attrition" accounts for non-native-like attainment rely on the assumption that the quality of input the heritage speakers receive resembles monolingual acquisition (in contrast to the quantity of input, which is severely reduced, as mentioned earlier), the "input quality approach" (cf. Rothman 2007; Pires & Rothman 2009; Pascual y Cabo & Rothman 2012) challenges the notion that the linguistic structures available to heritage speakers in their input always conform to the standard variety that monolinguals have access to. Thus, heritage speakers depend on input primarily from their parents. This input might already differ from monolingual varieties because of L1 attrition, cross-linguistic influence or general diachronic language change which might build on changes already extant in monolingual speech, but accelerated due to the lack of exposure to normative (written) standards (cf. Silva-Corvalán 1994). Following the "input quality approach", data from first generation immigrants (or any preceding generation) need to be taken into account when analyzing structures in heritage languages and establishing the variety that heritage speakers were exposed to. This is where research on L1 attrition and heritage linguistics complement each other.

It is for these reasons that the current volume focuses on issues concerning the role of L1 attrition in the speech of first generation immigrants and its implications for heritage language acquisition. Based on empirical evidence from a variety of heritage speaker contexts, the papers provide new insights into either attrition of the L1 among late sequential bilinguals and its impact on the proficiency of second generation heritage speakers or new insights into the role of different factors relating to quantity and quality of input in heritage language acquisition. The volume therefore contributes to the description and explanation of differences in the outcomes of heritage language acquisition and monolingual as well as foreign language acquisition. By taking into account different potential sources of variability in the development of heritage languages (e.g. cross-generational attrition, cross-linguistic influence in input, quantity of input, "incomplete acquisition", frequency and interface effects, sociolinguistic factors) the volume tries to disentangle these factors with regard to their impact on heritage language acquisition. As discussed in this introduction, all these factors can influence heritage language acquisition either in parallel or at different developmental stages. Furthermore, they can affect individual properties and structures or the heritage language as a whole (cf. Montrul 2008).

The individual contributions cover a wide range of typologically different heritage languages, including Romance (Spanish), Slavic (Russian) and Germanic languages (German, Norwegian) as well as Albanian, Turkish and Chinese as important immigrant languages in different European and North-American contexts. The presence of papers dedicated to the same heritage languages, but targeting heritage speaker communities in different countries (e.g. Turkish in the UK and Germany), as well as different heritage languages spoken in the same host community (e.g. Turkish and Russian as heritage languages in Germany) allows for cross-linguistic comparisons between different heritage language communities in the same linguistic environment and for comparisons regarding the effects of different majority languages on the same heritage language.

The volume starts with a set of three papers that directly compare the structural features under focus both in first generation immigrants (late sequential bilinguals) and heritage speakers. **Aalberse**, **Andringa**, **Faber** & **Lippe** investigate overt marking of definiteness on nouns referring to referents previously mentioned in the discourse among two generations of speakers of Wenzhounese Chinese in China and the Netherlands (parents and their children). Their data show an increase especially in the use of demonstrative constructions among second generation speakers of Wenzhounese Chinese, both in the homeland and in the Netherlands. Thus, the overuse of definiteness markers turns out to be the result of a generation rather than a location effect. Their results therefore point to the fact that the observed innovations cannot be solely due to language contact with Dutch (where overt definiteness marking is obligatory) or – in the case of heritage speakers in the Netherlands – to attrited input received from the first generation, but reflect more general internally motivated changes which might be accelerated in a heritage setting. However, the authors also consider the possibility of a heritage scenario for both second generation speakers in China and the Netherlands, as Wenzhounese is also spoken less often in China. Reduced input and use of Wenzhounese by second generation speakers could be a factor which affects definiteness marking in both locations and leads to parallel changes in the way definiteness is encoded (e.g. by preferring more explicit demonstrative constructions instead of neutralizing tone distinctions). Cross-generational data are also analyzed in the contribution of **Tuğba Karayayla**. By comparing data from both adult heritage speakers and first generation immigrants from Turkey in the UK to data from monolingual Turkish controls she explicitly puts to test the "input quality approach" described above. The feature under investigation is the marking of evidentiality in Turkish. Karayayla finds that the input the heritage speakers receive from the first generation immigrants regarding evidentials seems not to differ from monolingual input. However, the performance of heritage speakers in the application of indirect evidential structures clearly differs from that of monolinguals. In contrast to both the monolingual

controls and the late bilinguals (= first generation immigrants) they extend the use of direct evidentials also to non-witnessed contexts. Hence qualitatively different input which was provided by first generation immigrants cannot be the source of the phenomena attested in the data of the heritage speakers. To account for the findings among at least some heritage speakers, Karayayla looks at the effect of L1 input reductions over time on heritage speakers' accuracy in the use of indirect evidential structures in Turkish. The analysis reveals that less frequent L1 experience (input and output) before the age of 5 together with a richer L2 environment led to lower accuracy in evidentials by the respective heritage speakers. Karayayla thus concludes that the development of these heritage speakers was presumably never age-appropriate and they did not catch up with the more proficient heritage speakers at later stages, even if the presence of the L2 was weaker during their school years. Sufficient L1 experience in early childhood is thus crucial for the acquisition and long-term maintenance of evidentials in heritage Turkish. The third chapter by **David Giancaspro** compares heritage speakers of Spanish in the US to what he calls "Spanish-dominant controls", i.e. native speakers of Spanish who immigrated to the US after the age of 13 (= late bilinguals). Building on previous research concerning the tendency of heritage speakers of Spanish to produce fewer subjunctive mood forms by comparison with late bilinguals and/or monolinguals, Giancaspro shows that the replacement of lexically-selected subjunctive mood morphology by indicative forms is related to gaps in lexical rather than morphosyntactic knowledge. He conducts two experiments on the production and acceptability of lexically-selected subjunctive mood forms (or non-target indicative forms) following the complementizer *para que* in Spanish with both groups of Spanish-English bilinguals. The results obtained in both tasks reveal that despite a high general degree of accuracy with subjunctive mood, highly proficient heritage speakers are still significantly less accurate than Spanish-dominant controls, but only with lower frequency verbs. Thus, Giancaspro treats these results as evidence for the importance of the lexical frequency of verbs triggering intensional subjunctive mood with *para que* in Spanish. Given the reduced input that heritage speakers receive in their heritage language, they might fail to instantiate subjunctive mood features with verbs that occur infrequently in their Spanish input.

Frequency of structures is also a key topic that is investigated in the next set of five chapters. These focus on factors that shape heritage language acquisition, and most are more or less directly associated with the input that heritage speakers are exposed to. **Anderssen** & **Westergaard** investigate one specific aspect of word order in heritage Norwegian spoken in the US, namely the positioning of subjects and objects in relation to negation markers. Informationally given pronominal subjects and objects generally occur in front of the negation marker in Norwegian

(subject/object shift), although object shift is subjected to more structural constraints than subject shift. The authors examine the influence of two factors on the frequency of subject and object shift in heritage Norwegian: (a) structural similarity or difference in comparison to structures of the surrounding majority language, (b) frequency of occurrence. They use the Corpus of American-Norwegian Speech which contains data from 50 (mostly very old) heritage speakers of Norwegian who belong to the second, third or even fourth generation of Norwegian immigrants to the US. The presence of subject and object shift in this corpus is compared to corpora of monolingual Norwegian speech from the same dialectal background. The analysis reveals that both subject and object shift, despite their differing frequency of occurrence in monolingual Norwegian, are affected by restructuring in heritage Norwegian, thus indicating that frequency does not play a major role in the maintenance of these structures. Structural similarity between the heritage and surrounding majority language, however, seems to trigger changes in the heritage language: Complete structural overlap, i.e. when word order options are the same for both languages, leads to cross-linguistic influence regardless of the proficiency level in the heritage language. In this case, the word order which is preferred in English (but not in Norwegian) gets to be used more frequently by comparison with monolingual controls. **Jessica Diebowski**'s paper deals with another key variable for heritage language acquisition, namely the amount of use of the heritage language. She investigates data on the accuracy of gender assignment and gender agreement in Spanish and compares adult heritage speakers of Spanish living in the US (simultaneous bilinguals) to advanced English-speaking second language learners of Spanish. Her findings show that heritage speakers of Spanish perform at ceiling with regard to gender accuracy in written comprehension as well as oral production tasks, irrespective of the frequency of use of their heritage language. This stands in sharp contrast to adult L2 learners of Spanish, where the extent of exposure to and use of Spanish turned out to be a crucial factor for determining the success of gender acquisition in the L2. However, all informants were enrolled in Spanish-language classes (but not specific heritage speaker classes) which could account for the overall high accuracy of gender assignment and agreement in both groups and the equal distribution of heritage speakers irrespective of their amount of use of the heritage language.

A large number of factors and their impact on lexical proficiency in the heritage language is investigated in the paper of **Montanari**, **Abel**, **Tschudinovski** & **Graßer**. The authors look at effects of the amount of exposure to (= quantity of input) and use (= output) of the heritage language as well as socio-economic status and educational level of the parents and language dominance on the development of expressive and receptive vocabulary in children with Russian and Turkish as heritage

languages in Germany. One goal is to compare the two heritage speaker communities with regard to vocabulary size in the heritage language and the relevance of the above mentioned factors for lexical development in the heritage language. They tested overall 211 children (113 speakers of Russian and 98 speakers of Turkish) between the ages 6 and 10 by using a standardized picture naming task. The data reveal a good level of receptive vocabulary knowledge, but a limited expressive command of the test items. As usual in heritage language research, the authors point to a high level of interindividual variability. Both groups, however, displayed at best only a moderate development in lexical proficiency when the different age groups were compared. Furthermore, there is a systematic difference between the Russian and the Turkish-speaking groups which is accounted for by some of the social and pragmatic factors investigated in the study. A remarkable result of the comparison between the two heritage language communities is that whereas some of the investigated background variables yield the same effects for both groups (e.g. mother's proficiency in the heritage language), others show a diverging direction of impact on vocabulary knowledge (e.g. input patterns, institutional support of the heritage language or parents' highest level of education). The authors conclude that the degree of established networks (e.g. intense intergenerational contacts) and other social factors might contribute to the variability in outcomes of vocabulary acquisition among the two groups.

The diversity of sources of input in the heritage languages, especially access to written sources thanks to the availability of literacy skills, is the key factor under focus in the chapter by **Andreou**, **Dosi**, **Papadopoulo** & **Tsimpli**. They use a Sentence Repetition Task in order to explore the effects of biliteracy on the development of grammar and vocabulary in Albanian as a heritage language. Three different groups of children (aged 8–12) were investigated: (i) heritage speakers of Albanian living in Greece who do not receive institutional support in their heritage language and are thus monoliterate in the dominant language Greek, (ii) heritage speakers of Albanian living in Greece who receive written language support in Albanian (outside school) and are thus biliterate in Greek and Albanian, (iii) a control group of Albanian-Greek bilingual children who were born in Greece to Albanian-speaking families who later returned to Albania, but the children attended school in Greece and are therefore also biliterate. Data were collected in both languages, i.e. Greek and Albanian, and included elaborate questionnaires on literacy acquisition and language use. Besides the Sentence Repetition Task for Albanian and Greek, the children also carried out tasks targeting expressive vocabulary knowledge, non-verbal intelligence as well as verbal and non-verbal working memory. From a methodological point of view, the most important result of the study is that the children's (verbal and non-verbal) working memory abilities did

not predict their performance in the Sentence Repetition Task which underscores the suitability of this type of task for investigating linguistic proficiencies and not just memory skills of (bilingual) speakers. Factors like vocabulary knowledge or age (for L2 grammaticality score only) turned out to explain some variance in task performance. With regard to the grammaticality scores in Albanian, the three groups did not differ. However, differences were found between both heritage speaker groups and the group living in an Albanian environment regarding the ability to accurately repeat the sentences heard. The latter group outperformed both heritage speaker groups with regard to accuracy. The impact of L1 literacy (i.e. biliteracy in the heritage language context) on task performance turned out to be most prominent in the results of the verbal working memory task and the grammaticality scores in the L2 Greek (!). Here the heritage speaker group who received no support in the L1 Albanian scored significantly worse if compared to the other two groups. The authors treat this result as evidence for the validity of the Interdependence Hypothesis put forward by Cummins (see, e.g., Cummins 2001) which claims linguistic as well as cognitive benefits of bilingual children if they receive institutional support in both of their languages.

The starting point of **Elif Krause**'s paper is the well-known Interface Hypothesis (cf. Sorace & Serratrice 2009; Sorace 2011). She tests the prediction that structures involving interfaces between different cognitive domains are more problematic for bilingual language processing than structures that belong to one level only (e.g., syntax, morphosyntax, phonology, pragmatics etc.). The specific focus of the study is optional verb number marking in Turkish. Two different experiments are set up in order to check for effects of the semantics-morphosyntax and pragmatics-morphosyntax interfaces in producing the same structure under focus by Turkish heritage speakers in Germany. The use of plural markers on the verb in Turkish depends on semantic (animacy) as well as pragmatic (givenness) properties of the subject referents which makes them an ideal testing field for investigating interface effects. Krause uses grammaticality judgments by applying the Magnitude Estimation technique for data gathering. The results reveal that the sensitivity to animacy and givenness constraints in choosing overt plural marking on the verb is different for heritage speakers and age and education-matched monolingual controls. Thus, the heritage speakers applied a finer-grained analysis e.g. of animacy levels of the subject referents when deciding about the appropriateness of overt plural marking on the verb if compared to the monolinguals. However, this does not lead to an overrating of overt plural marking on the verb by the heritage speakers which would have been expected if cross-linguistic influence had been a relevant factor. The same tendency also applies to the pragmatic factor investigated in the second experiment, i.e. givenness of the plural subject. Here the bilinguals showed a finer distinction between different levels

of givenness than the monolinguals. Thus, heritage speakers behave differently if compared to monolinguals in being 'hyper-sensitive' with regard to single semantic and pragmatic properties of the plural subjects that constrain the use of overt plural marking on the verb. However, they were as accurate as the monolinguals in contexts where overt plural marking was categorically impossible in monolingual Turkish. This applies to both interface types.

The last series of chapters exclusively deals with aspects of L1 attrition, but takes into account the effects that L1 attrition can have on the acquisition of minority languages by subsequent generations. **Esther de Leeuw**'s paper looks at attrition effects in prosody in a group of L1 speakers of German living in Canada. More specifically, she investigates pitch level and pitch span in German and English by ten late sequential German-English bilinguals in comparison to a German and a Canadian English monolingual control group. The bilinguals emigrated from Germany to Canada as adults and had been living in the Vancouver area for an average of 40 years, yet they constitute classical candidates for L1 attrition. The data show that male German L1 speakers in Canada have on average a higher pitch level and a wider pitch span in both languages if compared to monolingual controls. Given the fact that pitch level is already higher in male monolingual speakers of German if compared to English, this finding is surprising as it contradicts the expected lowering of pitch levels in German by bilingual speakers due to the influence of the majority language English. De Leeuw suggests that this increase in pitch level differences is related to the social significance of pitch level alternations: According to some theories, a high pitch level is universally associated with friendly and non-aggressive behavior, whereas lower pitch levels characterize dominant and/or aggressive individuals. Bearing in mind the historically motivated low prestige of Germans in Canada as potential "enemies" associated with the Nazi regime and World War II, de Leeuw identifies the need to boost the image of the speakers as a possible motivation for pitch level raising, leading to non-monolingual like behavior in both languages. Social implications of pitch level and pitch span would thus be considered more important and desirable than the acquisition (L2 English) or maintenance (L1 German) of monolingual-like features of prosody. For women, however, both tendencies (convergence with the majority language and social indexing via prosody) go hand in hand, leading to higher pitch level and wider pitch span by comparison with L1 German norms and to an approximation (but still not native-like attainment) of the L2 English norms. De Leeuw concludes that her study shows that the political and social embedding of the bilingual community in the host country can also play a significant role in shaping outcomes of L1 attrition and L2 acquisition. The final chapter by **Shi Zhang** represents a pilot study on the potential attrition of perfective and durative

aspect marking in Mandarin spoken by Chinese immigrants who moved to the UK as literate adults and had lived there for an extended period of time (≥ 7 years) at the moment of data collection. Zhang looks at the interaction between lexical and grammatical aspect in Mandarin Chinese by investigating the acceptability of the perfective marker *le* and the durative marker *zhe* in combination with exponents of different lexical aspects. The data gained via an acceptability judgment task with 14 Mandarin-English bilingual speakers and 23 monolinguals from Mainland China showed no clear signs of L1 attrition in the domain of perfective and durative aspect marking on the part of the bilingual speakers. The author interprets these results as in accordance with the Interface Hypothesis, as aspect marking in Mandarin Chinese only involves an internal interface (syntax-lexicon) which proved to be less problematic for bilingual speakers by comparison with phenomena which involve external interfaces between syntax and other cognitive domains (cf. Sorace 2011).

The papers gathered in this volume provide ample evidence for the importance of distinguishing between the standard varieties that (most) monolingual speakers acquire in the homeland and the varieties that heritage speakers are exposed to in the host countries. However, the "input quality approach" does not necessarily account for all characteristics of heritage grammars and heritage lexicons. As shown in the papers of this volume, heritage language acquisition is shaped by a multitude of factors, including the social and political embedding of a heritage speaker community in the host community. This is what makes heritage languages a fascinating object of research, although we are just beginning to understand some of the basic mechanisms that shape their development.

References

Benmamoun, E., Montrul, S. & Polinsky, M. 2013. Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics* 39(3–4): 129–181.

Bylund, E. 2009. Maturational constraints and first language attrition. *Language Learning* 59(3): 687–715. https://doi.org/10.1111/j.1467-9922.2009.00521.x

Cummins, J. 2001. *Language, Power and Pedagogy: Bilingual Children in the Crossfire*. Clevedon: Multilingual Matters.

Dąbrowska, E. 2013. Heritage languages: A new laboratory for empirical linguistics. *Theoretical Linguistics* 39(3–4): 195–201.

Flores, C. M. 2010. The effect of age on language attrition: Evidence from bilingual returnees. *Bilingualism: Language and Cognition* 13(4): 533–546.
https://doi.org/10.1017/S136672890999054X

Flores, C. M. 2012. Differential effects of language attrition in the domains of verb placement and object expression. *Bilingualism. Language and Cognition* 15(3): 550–567.
https://doi.org/10.1017/S1366728911000666

Köpke, B. & Schmid, M. 2004. Language attrition. The next phase. In *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues* [Studies in Bilingualism 28], M. S. Schmid, B. Köpke, M. Keijzer & L. Weilemar (eds.), 1–43. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.28.02kop

Kupisch, T. 2013. A new term for a better distinction? A view from the higher end of the proficiency scale. *Theoretical Linguistics* 39(3–4): 203–214.

Kupisch, T. & Rothman, J. 2018. Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism* 22(5): 564–582.

Montrul, S. 2008. *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor* [Studies in Bilingualism 39]. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.39

Montrul, S. 2016. *The Acquisition of Heritage Languages*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139030502

Pallier, C. 2007. Critical periods in language acquisition and language attrition. In *Language Attrition: Theoretical Perspectives* [Studies in Bilingualism 33], B. Köpke, M. S. Schmid, M. Keijzer & S. Dostert (eds.), 155–168. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.33.11pal

Pascual y Cabo, D. & Rothman, J. 2012. The (Il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied Linguistics* 33(4): 450–455. https://doi.org/10.1093/applin/ams037

Pavlenko, A. 2004. L2 influence and L1 attrition in adult bilingualism. In *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues* [Studies in Bilingualism 28], M. S. Schmid, B. Köpke, M. Keijzer & L. Weilemar (eds.), 47–59. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.28.04pav

Pires, A. & Rothman, J. 2009. Disentangling sources of incomplete acquisition: An explanation for competence divergence across heritage grammars. *International Journal of Bilingualism* 13(2): 211–238. https://doi.org/10.1177/1367006909339806

Polinsky, M. 2006. Incomplete acquisition: American Russian. *Journal of Slavic Linguistics* 14(2): 191–262.

Polinsky, M. 2008. Russian gender under incomplete acquisition. *The Heritage Language Journal* 6(1): 40–70.

Polinsky, M. 2018. *Heritage Languages and Their Speakers*. Cambridge: CUP.

Polinsky, M. & Kagan, O. 2007. Heritage languages: In the 'wild' and in the classroom. *Language and Linguistics Compass* 1(5): 365–395. https://doi.org/10.1111/j.1749-818X.2007.00022.x

Rothman, J. 2007. Heritage speaker competence differences, language change, and input type: Inflected infinitives in Heritage Brazilian Portuguese. *International Journal of Bilingualism* 11(4): 359–389. https://doi.org/10.1177/1367006907011004201

Rothman, J. 2009. Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism* 13(2): 155–163. https://doi.org/10.1177/1367006909339814

Rothman, J. & Treffers-Daller, J. 2014. A prolegomenon to the construct of the native speaker: Heritage speaker bilinguals are natives too! *Applied Linguistics* 35(1): 93–98. https://doi.org/10.1093/applin/amt049

Schmid, M. S. 2011. *Language Attrition*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511852046

Schmid, M. S. 2013. First language attrition. *Linguistic Approaches to Bilingualism* 3(1): 96–117. https://doi.org/10.1075/lab.3.1.05sch

Schmid, M. S. & Köpke, B. 2007. Bilingualism and attrition. In *Language Attrition: Theoretical Perspectives* [Studies in Bilingualism 33], B. Köpke, M. S. Schmid, M. Keijzer & S. Dostert (eds), 1–7. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.33.02sch

Schmid, M. S. & Köpke, B. 2017. The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism* 7(6): 637–667. https://doi.org/10.1075/lab.17058.sch

Silva-Corvalán, C. 1994. *Language Contact and Change: Spanish in Los Angeles*. Oxford: Clarendon Press.

Sorace, A. 2011. Pinning down the concept of 'interface' in bilingualism. *Linguistic Approaches to Bilingualism* 1(1): 1–33. https://doi.org/10.1075/lab.1.1.01sor

Sorace, A. & Serratrice, L. 2009. Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism* 13(2): 195–210. https://doi.org/10.1177/1367006909339810

Definiteness in Wenzhounese Chinese in the Netherlands and in China

Evidence for generational change in two locations

Suzanne Aalberse, Sible Andringa, Martina Faber and Phine Lippe
University of Amsterdam

By comparing two generations of speakers in China and the Netherlands, we investigated whether Wenzhounese Chinese as spoken by heritage speakers in the Netherlands might be subject to change due to its contact with the Dutch language. To this end, we considered how nouns referring to already mentioned referents were encoded, hypothesizing an increase in overtly marked already mentioned referents in the speech of second generation speakers in the Netherlands. Also, shifts in the use of classifier and demonstrative constructions were investigated as carriers of definite meanings. The data showed clear generation effects, especially in the use of classifier and demonstrative constructions, but no location effects. Although the interpretation of the data is compromised by the fact that Wenzhounese as spoken China is under threat, the study suggests that innovations found in heritage speakers may not necessarily be related to the heritage scenario, but can also reflect more general contact effects or internally motivated changes.

Keywords: cross-linguistic influence, heritage speakers, definiteness, Wenzhounese Chinese, change across the lifespan, co-activation

1. Introduction

If heritage speakers speak differently than their peers in the home country, there are various possible explanations. A possible explanation is that heritage speakers have less input of the language and therefore do not acquire those aspects of the language that are dependent on frequent input (cf. O'Grady, Kwak, Lee & Lee 2011). Another explanation is that heritage speakers have received input that is different from the input homeland speakers received because their parents' speech showed

traces of language contact, and/or because the heritage language is used in different contexts than in the homeland (Pires & Rothman 2009). A third possibility is that differences are due to (accelerated) internally motivated language change.

This study aims to disentangle sources of change by comparing two generations of speakers of Wenzhounese Chinese in the Netherlands and in China. We investigated second generation speakers and one of their parents. If change is caused by language contact, we should observe this change in both generations in the Netherlands, e.g. we would expect a location effect. If change is motivated by internal factors, we would expect change in all second generation speakers independent of location; e.g., we would expect a generation effect. If change is driven by the heritage scenario, we would expect an interaction effect of generation and location, such that only second generation speakers in the Netherlands would show the change. The specific phenomenon investigated was the encoding of already mentioned referents, and we specifically looked for the occurrence of shifts in frequency, which has been coined "indirect transfer" by Silva-Corvalán (1994: 4) and "frequential copying" by Johanson (2002: 292). Dutch and Wenzhounese Chinese partly differ in how they encode already mentioned referents, but there is also overlap. Partial overlap can facilitate cross-linguistic influence. Section 2 provides more information on change in case of partial overlap and on the phenomenon under investigation, namely definiteness. Section 3 provides the methodology, Section 4 the results and Section 5 the conclusion.

2. Theoretical framework

2.1 Cross-linguistic influence and change across the lifespan

Jarvis and Pavlenko (2008: 1) define cross-linguistic influence as "the influence of a person's knowledge of one language on that person's knowledge or use of another language". It is often hypothesized that the occurrence of cross-linguistic effects is related to some degree of overlap between the languages. Jansen, Lalleman and Muysken (1981), for example, formulated the alternation hypothesis that states that bilinguals prefer a construction that is shared by both of their languages at the expense of a construction that is disallowed in one of their languages. Similarly, Sánchez (2004) formulates the functional convergence hypothesis which states that syntactic convergence is most likely to occur in bilingual speakers when a set of features associated with a functional category is partially divergent and partially similar. Both hypotheses expect a shift in frequency of use and presuppose that bilingual speakers can subconsciously make connections between elements or constructions of two or more languages. Such connections have been referred to

as "interlingual identification" by Weinreich (1979) or "equivalence relations" by Heine and Kuteva (2005).

Schmid (2011) and Moro (2016) found that people who use both of their languages actively are the greatest innovators and related language use patterns during the lifespan (and not the level of proficiency) to the likelihood that grammars converge. Converging grammars are probably cognitively easier to use because the same structure can be applied to both languages. The observation that language use patterns (e.g., the frequent use of two languages in similar settings) correlate with convergence can be explained in two (non-exclusive) ways: (1) frequent co-activation might facilitate interlingual identification and (2) the benefit of convergence may be greater to those speakers who frequently co-activate their two languages.

2.2   Definiteness marking, language contact and bilingual acquisition

This subsection introduces the main concepts concerning definiteness marking. All languages can encode if a referent is unique in a given discourse, because it was already mentioned or presupposed in the preceding discourse. Such unique referents are referred to as "definite" (cf. Heim 1991; Ionin, Ko & Wexler 2004: 5; Cho & Slabakova 2014: 161). Languages differ with respect to how they mark definiteness and they can differ in the extent to which such marking is obligatory.

Some languages like English or Dutch obligatorily encode definiteness on nouns via dedicated morphology. The article *the*, for example, encodes definiteness and no other grammatical category. The category of markers that have definiteness marking as their primary function are referred to as simple definites by Lyons (1999) and as direct markers by Cho & Slabakova (2014). Morphemes such as demonstratives or possessive markers that are inherently definite and encode definiteness indirectly, but that mainly encode some other grammatical category (such as deictic information), are referred to as complex definites by Lyons (1999) and as indirect markers by Cho and Slabakova (2014).

Apart from differing in the obligatory nature of marking, languages may also differ in the types of definites they possess: they either have complex definites only or both simplex and complex definites. Moro (2016) suggests that a grammatical category that is obligatorily encoded in one language and optionally encoded in the other language, is likely to undergo an increase in overt marking in the language where the category is optionally marked. Cho and Slabakova (2014) showed that not only the extent of the overlap between the languages, but also the type of encoding affected the learning process. Direct and morphological marking seemed easier to acquire than indirect and/or non-morphological marking of definiteness (for example via word order).

There are various traditions in linguistics which suggest that the encoding of definiteness is susceptible to cross-linguistic influence. Long term language contact studies show that languages without articles tend to develop articles when in contact with a language with articles. Backus, Doğruöz and Heine (2011) provide an extensive list of such examples. New definite articles are most likely to be etymologically derived from demonstrative pronouns, but other pathways are possible, too. Not only language contact studies suggest that definiteness is susceptible to cross-linguistic influence. Several acquisition studies report on cross-linguistic influence in the domain of definiteness marking. An important observation in these studies is that structural similarity between languages in the domain of definiteness appears to facilitate learning. Montrul and Ionin (2012) showed that L2 learners of Spanish and heritage speakers of Spanish in the US hardly made any mistakes in the use of articles; mistakes were seen only in those domains where the languages differ. In contrast to learners with an L1 with articles, learners with an L1 without articles have long and persistent problems in learning articles. These problems are widely documented (see, for example, Kharma 1981; Huebner 1983; Agnihotri, Khanna & Mukherjee 1984; Master 1987; Thomas 1989; Butler 2002; Ionin, Ko & Wexler 2004; Goad & White 2008; Mayo & Hawkins 2009, among many others). Various types of errors have been found in the production of such learners. They have shown omission errors where nouns that need an article are left bare; they have mixed up definite and indefinite articles; they have overextended articles to places that do not need them; and they have used phonologically stronger and more lexicalized forms instead of the article (e.g., the use of *that* instead of *the* and the use of *one* instead of *a*).

2.3   Definiteness marking in heritage languages

When the socially dominant language marks definiteness obligatorily, while the heritage language does not (as is also the case in the present study), overextension of the use of overt marking in the heritage language appears to be a likely result. This has been observed in three studies. Polinsky (2006) described the characteristics of heritage Russian in the US. One of the features she observed was the extensive use of demonstratives in narratives. An example of such a narrative (taken from Polinsky 2006: 247) is shown in (1). Both the noun *boy* and the borrowed noun *policeman* in (1) are preceded by the demonstrative *ètot* ('this'). Polinsky (2006) reports that such use of demonstratives is perceived as redundant by Russian speakers in Russia.

(1)   Ètot mal' čik$_i$ togda ubežal   on$_i$ ubežal i   vot ètot policeman
      this boy   then   ran away he ran away and DM this policeman

The frequency of demonstratives in heritage Russian appeared much higher than in homeland Russian. Polinsky (2006) provided two possible explanations for the overextension of demonstratives. Either heritage speakers want to avoid ambiguity or they compensate for the lack of articles in Russian.

The second study on the use of definiteness markers in heritage speakers is Moro (2016). She described the use of both the enclitic -*nya*, indicating definiteness and possession, and the demonstrative *itu* in three groups of speakers: homeland speakers, first generation Ambon Malay migrants and heritage Ambon Malay speakers in the Netherlands. She found that speakers in the Netherlands used more markers to encode definiteness of already mentioned nouns than speakers in Ambon. She observed this both in first and second generation speakers in the Netherlands. The specific marker that the speakers used varied: some overused the demonstrative pronoun *itu*, some overused the possessive and definite affix -*nya* and some used a mix of both. The effects were strongest in speakers who grew up as sequential bilinguals and who lived outside Moluccan wards (special neighborhoods for Moluccans in the Netherlands). Moro did not find a significant difference between generations. Since her data reported on a limited number of speakers, the absence of a significant effect between generations may have been a problem of statistical power. Moro suggested that the overextension patterns observed were due to cross-linguistic activation. Her speakers were quite fluent in Ambon Malay and they used Dutch frequently. In such situations of active use of both languages, more syntactic priming may have made convergence of the languages more likely.

The third study that observed extended use of definiteness markers in a heritage context is Aalberse, Zou & Andringa (2017). This study compared two generations of speakers of Mandarin Chinese in the Netherlands to one generation of speakers in mainland China. The study reported that all speakers in the Netherlands (first and second generation) were less likely to use bare forms when referring to already mentioned nouns than homeland speakers. When an already mentioned referent was referred to, the noun referring to this referent is frequently modified by a demonstrative.

The results by Polinsky (2006), Moro (2016) and Aalberse et al. (2017) can all be interpreted as in line with the alternation hypothesis as well as with the functional convergence hypothesis. The reduced reliance on bare nouns indicates that the construction with overlap in the two languages (i.e. overt and morphologically marked encoding of definiteness) is preferred at the expense of a construction that is disallowed in one of the languages, which is the bare form. This is in line with the alternation hypothesis. One could also assume that this shift in frequency is the result of a recategorization of feature bundles associated with demonstratives. The demonstrative marker, for example, might have become less sensitive to the 'deictic' marking feature, making it more appropriate in a wider range of contexts and

thus enhancing its frequency. At this moment we cannot decide which of the two theories actually accounts for the overextension of definiteness marking. Polinsky (2006) has suggested that explicitness could also cause the observed shifts: Speakers might just want to be more explicit, which is a strategy that is associated with insecurity in non-fluent speakers. However, since Moro finds that not the level of fluency, but frequent co-activation is the best predictor of overt definiteness marking in her participants, it seems that explicitness alone cannot be the complete explanation.

2.4   Definiteness marking on already mentioned nouns in Wenzhounese Chinese and Dutch

Languages can differ in the extent to which definiteness is marked obligatorily and in whether they use a direct or indirect marking strategy. In the present section, Dutch and Wenzhounese are compared for these features.

Wenzhounese Chinese is spoken in the city of Wenzhou in the province of Zhejiang and is a southern Wu-language. Wu-languages show large internal variation, typically to the point of mutual unintelligibility (Cheng & Sybesma 2005: 259). Like other Chinese languages, Wenzhounese can use classifier noun combinations to encode new information as shown in (2a). Unlike other Chinese languages, Wenzhounese can encode definiteness by neutralizing tone (Cheng & Sybesma 2005) on the classifier used. An example of tone neutralization of a classifier to indicate definiteness (taken from Cheng & Sybesma 2005: 268) is provided in (2). The example gives a minimal pair where the classifier *paŋ* with its original third tone indicates indefiniteness whereas it indicates definiteness if the neutralized tone version (tone 7 in b) is used. Using tone on a classifier is one option to encode definiteness. Apart from using tone-neutralized classifiers, definite nouns can be left bare or can be encoded with complex definite elements such as possessives or demonstratives.

(2)   Minimal pair paŋ$^{3/}$ paŋ^7
    a.   ŋ^4 ɕi3   ma^4 paŋ^3 si^1 iɕŋ
          I   want buy CL   volume book
          "I want to buy a book."
    b.   ŋ^4 ɕi3   ma^4 paŋ^7 si^1 iɕŋ
          I   want buy CL   volume book
          "I want to buy the book."

The main difference between definiteness marking in Dutch versus Wenzhounese is obligatoriness: Wenzhounese can, but does not have to encode definiteness,

whereas Dutch obligatorily encodes definiteness either with an article or with a complex definite marker such as a possessive marker or a demonstrative. The other difference is that Dutch encodes definiteness with a separate morpheme, an article, whereas Wenzhounese modifies a classifier and thus encodes definiteness via a non-segmental strategy.

2.5   Present study

Given that definiteness is open to cross-linguistic influence, we investigated data collected for the ERC Traces of Contact project (ERC #230310) (PI Pieter Muysken) for Wenzhounese on the presence of overt marking of definiteness in encoding already mentioned referents. We hypothesized that the obligatory nature of definiteness marking in Dutch could trigger an increase in overt marking of definiteness in the Wenzhounese speakers in the Netherlands. By comparing two generations of speakers in China and the Netherlands, we investigated whether changes in the distribution of bare versus overtly encoded nouns were present. We also looked at shifts in the use of classifier and demonstrative constructions as candidates for overtly encoding definiteness in Wenzhounese.

3.   Methodology

3.1   Participants

For this empirical study we tested 5 Wenzhounese Chinese families living in the Netherlands (first and second generation) and five Wenzhounese Chinese families living in Wenzhou. One parent and one (young) adult child from each family were included in the analysis yielding a total of 20 speakers. This resulted in four sub-groups: the first generation (age: range 42–48; mean: 45.2) in China, the first generation in the Netherlands (age: range 38–56; mean: 47.0), the second generation (age: range 20–28; mean: 22.2) in China and the second generation in the Netherlands (age: range 17–32; mean: 23.0). The participants were recruited by Xiaoli Dong in or near the Arnhem area in the Netherlands and by Danhong Wa in Wenzhou, China through the ERC Traces of Contact project. The first generation speakers in the Netherlands are Wenzhounese Chinese native speakers who were born in Wenzhou. They moved to the Netherlands to work and have been staying there since then. All second generation speakers in the Netherlands were born in the Netherlands.

3.2   Traces of Contact corpus

The elicitation material was part of the CORE-elicitation kit of the Traces of Contact project (ERC project #230310). The CORE- elicitation consisted of three parts: (1) a set of short clips and pictures that participants had to describe after seeing; (2) a set of videos that speakers had to describe while seeing and (3) an oral sociolinguistic interview. This study reports on how speakers encode already mentioned referents while describing the second part of the elicitation kit, namely the video clips. The video material was most suitable for investigating already mentioned referents because the same referents kept returning several times during the video triggering the use of definiteness marking. The video elicitation task consisted of eight videos from the German cartoon series "Die Sendung mit der Maus", three videos were created by Sotaro Kita et al. at the MPI (Kita 1995) and three videos were created by Geoffrey Haig and Stefan Schnell at the University of Kiel (Haig & Schnell 2010). Some characters like the mouse and the elephant (as illustrated below) recurred in several videos. The average length of the videos was about half a minute (shortest video: nine seconds, longest one: 56 seconds) and they were presented in three randomly generated orders. The participants were assigned to one of these orders and they were asked to describe what they saw on the laptop in front of them while watching.
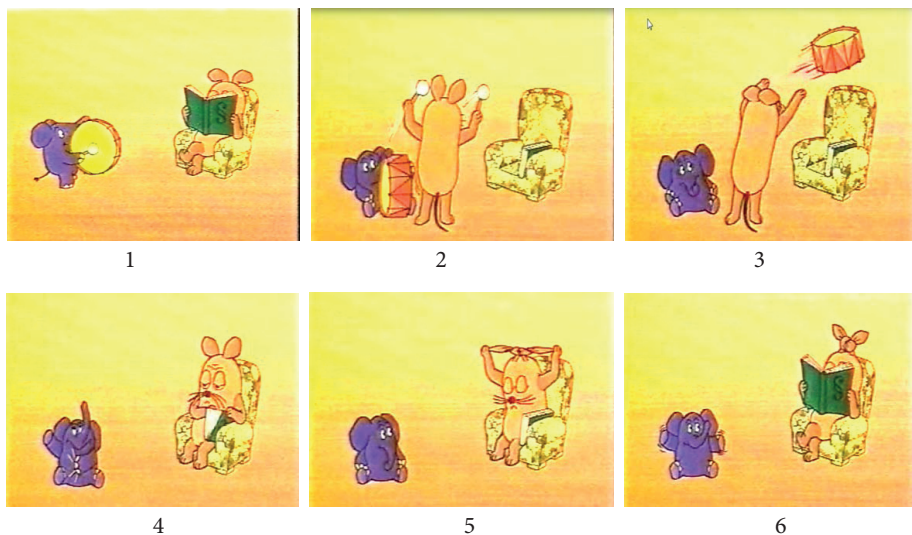
Figure 1. Example of a video clip employed for data collection

3.3   Coding

The ELAN[1] program (Sloetjes & Wittenburg 2008) was used to code and analyze the data. All nominal constituents in the data were coded in one independent tier as *first mention* if it was introduced by the participant for the first time, or as *already mentioned* if the referent had been introduced before. When characters like the mouse and the elephant reappeared in a video, it was not always clear if the participant recognized these characters as newly mentioned information. If in these cases the participant used a bare form or simply attached a demonstrative to the noun, they were coded as unclear; if the participant used a numeral *one* (and a classifier) or clearly stated that it was the same referent, then it was coded as *first mention* or *already mentioned* respectively. Among all the already mentioned referents, those that showed up as bare forms were marked *bare nouns* as shown in (3), nouns preceded by classifiers only as shown in (4) as *classifier,* nouns preceded by demonstratives such as (5) and (6) were marked in another independent tier as *demonstratives*. Possessive pronouns and other elements that made nouns inherently definite and thus not open to more definite marking were encoded as *other*.

(3)   bare noun
       lets'ei- mu- ?yi- mu
       老鼠　摸　一　摸
       mouse- touch-one-touch
       "The mouse touched (the cake)."

(4)   classifier noun
       kai- lets'ei,   hong-hong
       个　老鼠，　嗅嗅
       CL- mouse- smell-smell
       "The mouse smelt (the cake)."

(5)   demonstrative noun
       ki- meimei- jjau- xia-ha-va
       居　姆姆　　球　吸-丐-还
       this-child-   ball- throw-to-return
       "This child threw the ball back."

(6)   demonstrative and classifier
       hi- kai- seing- pai-   jie- tang
       许　个　细儿　背　　张　凳
       that- CL- child- carry- CL- chair
       "The child carries a chair."

---

1.   ELAN is a professional tool for the creation of complex annotations of video and audio resources designed by the Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. <http://tla.mpi.nl/tools/tla-tools/elan/>

3.4 Data analysis

The analyses presented here were run on already mentioned contexts only. Generalized linear mixed-effects logistic regression modeling was used to investigate if the way nouns referring to already mentioned referents were marked depended on location (China or the Netherlands) and generation (first or second). This technique lends itself particularly well for the analysis of hierarchically organized data, as is the case here, where language use samples are nested within participants, who themselves are nested within families. In this technique, parameter estimates are computed for both random and fixed effects. Random effects parameters are estimates of the variance that is associated with a random effect (participants and families in this case). According to common recommendation, we assessed the decrease in the deviance statistic to assess model improvement for random parameters. Fixed effects estimates must be interpreted as in logistic regression; they are log odds ratios or logits. In this study, they express the chance that a particular type of realization is more likely to occur at a certain location or within a certain generation. We always ran fully specified generalized linear mixed-effects models, which in practice meant that the random structure consisted of participants nested in families (Barr, Levy, Scheepers & Tily 2013). Random slopes for location and generation could not be included as these are between participant and family variables. All analyses were run in R (R core team 2015) using the lme4 package (Bates, Maechler, Bolker & Walker 2015).

4. Results

In the analyses presented here we investigated the likelihood of the occurrence of overtly marked nouns and whether this occurrence depended on generation and location. Table 1 provides descriptive statistics per speaker. As can be seen, the 20 speakers produced a total of 1541 contexts relevant for the analysis, i.e., contexts where the referent had already been mentioned in the previous discourse. In 684 instances, speakers produced some kind of overt marking, which is 44% of the total number of occurrences. In 424 contexts speakers used a classifier construction, while a demonstrative construction was used in 221 contexts. The table also shows that there were considerable individual differences in the use of marked nouns and the use of classifiers and demonstratives; patterns are not easily detectable without inferential statistics, but it does seem that classifiers were used more by first generation speakers, while demonstratives were preferred by second generation speakers. Also note the first generation speaker of family 4, who seemed to use an exceptionally high number of overtly marked nouns in comparison to other speakers. Table 2 summarizes the same data, organized by location and generation.

Table 1. Occurrence of already mentioned nouns, overt marking and classifier and demonstrative constructions per speaker

Family	Location	Generation	Total already mentioned nouns	Overtly marked (% of total)	Classifiers (% of overtly marked)	Demonstratives (% of overtly marked)
1	Netherlands	first	115	38 (33%)	30 (79%)	4 (11%)
	Netherlands	second	139	79 (57%)	55 (70%)	19 (24%)
2	Netherlands	first	53	21 (40%)	18 (86%)	0 (0%)
	Netherlands	second	105	37 (35%)	14 (38%)	20 (54%)
3	Netherlands	first	94	15 (16%)	13 (87%)	2 (13%)
	Netherlands	second	96	67 (70%)	31 (46%)	34 (51%)
4	Netherlands	first	115	93 (81%)	70 (75%)	15 (16%)
	Netherlands	second	91	54 (59%)	13 (24%)	38 (70%)
5	Netherlands	first	62	22 (35%)	17 (77%)	3 (14%)
	Netherlands	second	56	23 (41%)	15 (65%)	7 (30%)
6	China	first	40	5 (13%)	0 (0%)	5 (100%)
	China	second	73	34 (47%)	4 (12%)	29 (85%)
7	China	first	33	11 (33%)	10 (91%)	0 (0%)
	China	second	60	12 (20%)	6 (50%)	6 (50%)
8	China	first	62	22 (35%)	18 (82%)	2 (9%)
	China	second	103	44 (43%)	37 (84%)	7 (16%)
9	China	first	39	9 (23%)	8 (89%)	1 (11%)
	China	second	68	28 (41%)	16 (57%)	11 (39%)
10	China	first	75	37 (49%)	33 (89%)	4 (11%)
	China	second	62	33 (53%)	16 (48%)	14 (42%)
Total			1541	684 (44%)	424 (62%)	221 (32%)

Table 2. Summary of data by location and generation

Speaker	Overtly marked (% of total)		Classifiers (% of overtly marked)		Demonstratives (% of overtly marked)	
	China	Netherlands	China	Netherlands	China	Netherlands
First generation	33.7%	43.1%	85.2%	86.0%	14.8%	14.0%
Second generation	41.3%	53.4%	50.0%	52.0%	45.9%	48.0%

It was hypothesized that change might be visible in a shift towards overt encoding of definiteness in Wenzhounese. Therefore, the first step was to investigate in the total set of already mentioned contexts ($n = 1541$) if the speakers' choice to refer to already mentioned referents with overtly modified nouns was related to location (the Netherlands or China) and generation (first or second). For this, we

ran a model that included the random structure of participants nested in families, as well as location and generation as fixed effects. The analysis showed that speakers used significantly more bare forms than marked forms (857 vs. 684; $b = -0.382$, $SE = 0.158$, $p = 0.016$, OR = 0.68). Table 2 shows a tendency that more overtly marked nouns were used by second generation speakers and speakers in the Netherlands. Second generation speakers used overtly marked nouns 1.61 times for each marked noun used by first generation speakers, but this effect was not significant ($b = 0.474$, $SE = 0.316$, $p = 0.134$, OR = 1.61). Similarly, speakers in the Netherlands were 1.63 times more likely to use marked forms, but this effect was also not significant ($b = 0.490$, $SE = 0.316$, $p = 0.121$, OR = 1.61). The interaction between location and generation was not significant, too ($b = 0.038$, $SE = 0.632$, $p = 0.952$, OR = 1.04). If the first generation speaker of family four is excluded (see the discussion section for why this might be appropriate), then a significant generation effect emerges ($b = 0.698$, $SE = 0.242$, $p = 0.004$, OR = 2.01), but no location effect or generation by location interaction. To investigate the effect of family, we tested if significant portions of the variance could be attributed to differences between families by checking if model fit improved when participants were modelled as nested within families (as compared to a baseline model that included the by-participant random effect only). We found no reduction in the deviance statistic, which expresses how much variation a model explains ($\chi^2(1) = 0$; p>0.05). Conversely, we did find that the inclusion of the random by-participant intercept lead to significantly better model fit ($\chi^2(1) = 54.996$; $p < 0.001$) when compared to a model that included family only.

We then explored whether there was evidence of differentiated use of classifier and demonstrative constructions according to location or generation, which might also be indicative of change. For this analysis, we considered the set of marked already mentioned referents only and included only instances of classifier and demonstrative constructions, meaning that 39 instances were excluded where other forms of overt marking such as possessive constructions were used. The dependent variable expressed whether a classifier or demonstrative was used. The analysis showed that demonstratives were used less frequently than classifiers (424 vs. 221; $b = -1.653$, $SE = 0.562$, $p = 0.003$, OR = 0.19). The results also showed that second generation speakers were 4.86 times more likely to use demonstrative forms than first generation speakers ($b = 1.580$, $SE = 0.731$, $p = 0.031$, OR = 4.86) and therefore less likely to use classifiers, which was also clearly suggested by Table 2. The effects of location ($b = -0.504$, $SE = 0.764$, $p = 0.509$, OR = 0.604) and the generation by location interaction ($b = 0.521$, $SE = 1.001$, $p = 0.52$, OR = 1.683) were not significant, despite fairly large odds ratios. We again tested for a family effect by investigating the random effects structure in this dataset and found a similar pattern as before. The inclusion of family as a random variable in which participants were nested did

not improve model fit ($\chi^2(1) = 0$; p>0.05) in comparison to a model that included the random participant effect only, while the inclusion of the by-participant intercept significantly improved model fit ($\chi^2(1) = 17.257$; $p < 0.001$).

5.   Discussion

The present study did not find evidence of differentiated use of overt marking – or differentiated use of classifiers and demonstratives within overtly marked nouns – according to location, and therefore does not present evidence of contact-induced change. This is different from Aalberse, Zou and Andringa (2017) and Moro (2016), which both did find evidence for a location effect indicative of a shift due to language contact. There may be several reasons for the absence of location effects in this study. Of course, it is possible that Wenzhounese as spoken in the Netherlands is not affected by the close proximity of Dutch, at least not in how definiteness is expressed. Possibly this is due to the fact that – unlike in the case of Malay, Russian and Mandarin Chinese – Wenzhounese can encode definiteness directly already. Another reason could be that our analyses were insufficiently sensitive to detect increased tendencies of expressing definiteness in Wenzhounese spoken in the Netherlands. One reason for this insufficient sensitivity may have been the small number of participants. There were only ten speakers in the Netherlands, and no more than five speakers who grew up with Wenzhounese there. We consider this explanation likely, as the odds ratios observed for the location effects were fairly large. If we consider the group percentages presented in Table 2, then we see that speakers in the Netherlands always used more overtly marked forms. The location effects probably failed to reach significance because of the small number of speakers and the substantial individual variation.

Although the odds ratio suggests that second generation speakers used overtly marked nouns 1.61 more than first generation speakers, this difference (that was similar to the data reported in Moro 2016) did not reach significance. However, there were clear generation effects for the use of classifiers and demonstratives, which seems to have shifted from classifiers towards demonstratives in the second generation. In addition, it should be noted that the absence of a generation effect for overt marking seemed to be caused by the somewhat exceptional behavior of one first generation speaker in the Netherlands, who used a high number of overtly marked nouns. Without this speaker, second generation speakers in both China and the Netherlands were actually twice as likely to use overtly marked forms. This parent was the only person in our Dutch dataset who reported using Wenzhounese outside the family context frequently, meaning that context of use of Dutch and Wenzhounese frequently overlapped for this participant. Perhaps

the frequent co-activation of the two languages accelerated an internal change in Wenzhounese in this particular first generation speaker. Although this cannot be concluded from only one participant, this suggests that the true driving force of contact-induced change is possibly the level of co-activation of both heritage and host language. Chances of finding high levels of co-activation may be higher in second generation speakers, but they may of course also occur in first generation speakers. This would mean that generation is really a proxy of co-activation, and co-activation the driver of change. It would be useful the replicate the present findings and include more detailed information about the use of both languages to determine the degree of conjoint use.

Finally, the generation effect observed is consistent with the possibility that Wenzhounese as spoken in the Netherlands and China are changing independently, but in a similar direction either because of internally motivated change or because the heritage scenario is at play in both locations. Wenzhounese is spoken less both in the Netherlands and in China and this reduced use could have repercussions for definiteness marking. One possible implication is that reduction in use could affect the ability to produce and perhaps perceive tone. Since definiteness is encoded by tone neutralization in Wenzhounese, the loss of the ability to distinguish tone could make the neutralized classifier construction vulnerable. Speakers might no longer be able to adequately distinguish non-neutralized from neutralized tones – the only factor that separates definite nouns from indefinite nouns. Alternatively, non-segmental encoding of definiteness might be vulnerable more generally (cf. Cheng & Sybesma 2005) and might therefore be instable in situations of intense language contact.

The generation effect was strongly visible in a shift from a preference to use classifiers in definite contexts by first generation speakers, to the use of demonstratives by second generation speakers in both China and the Netherlands. This strong generation effect in formal encoding – second generation speakers were 4.8 times more likely to use demonstratives – was not something we expected. Perhaps, if it is true that tone is losing its function to encode definiteness, second generation speakers are adopting demonstrative constructions as an alternative form for encoding definite meanings. Polinsky (2006) has argued that speakers may revert to more explicit forms in the face of uncertainty. Such a strategy may be at play here as well. Uncertainty about tone or uncertainty about language proficiency in general may have driven second generation speakers to use more explicit forms. Demonstrative constructions are arguably more explicit forms for encoding definiteness than classifiers, both because demonstratives encode more information than classifiers and because demonstratives are segmental carriers of definiteness and therefore more robust in cases of disruption than non-segmental carriers of definiteness such as classifiers. More research is needed that includes an analysis of tone changes and an analysis of language proficiency more generally to get to the bottom of this.

6.   Conclusion

By comparing two generations of speakers in the Netherlands and China, we found a change in second generation speakers, similar in both locations, namely the use of demonstratives at the expense of classifiers to encode definiteness. We could be witnessing internally motivated change or an effect of the weakened use of Wenzhounese in China as well as in the Netherlands. While the generation effect was clear, the expected location effect was not. There may have been several reasons for this, one being that the changes present in Wenzhounese in the Netherlands also occur in China because the language is threatened. The other explanation is a lack of power, which was supported by substantial effect sizes associated with location effects.

This study illustrates that language change is not easily demonstrated, and observed effects are not easily interpreted. This study was limited by the unclear role that tone may have played and by the relatively small number of speakers included. A strength of this study was that transnational and cross-generational data were collected. Although a bit problematic for the Wenzhounese situation where the home language is under threat, more generally including two generations in the homeland potentially allows one to determine whether shifts observed in heritage speakers are due to the language contact situation or to internally motivated change. The method can and should be improved by including sensitive measures of language use and co-activation.

References

Aalberse, S., Zou, Y. & Andringa, S. 2017. Extended use of demonstrative pronouns in two generations of Mandarin Chinese speakers in the Netherlands. In *Cross-linguistic Influence in Bilingualism: In Honor of Aafke Hulk* [Studies in Bilingualism 52], E. Blom, L. Cornips & J. Schaeffer (eds), 25–48. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.52.03aal

Agnihotri, R. K., Khanna, A. L. & Mukherjee, A. 1984. The use of articles in Indian English: Errors and pedagogical implications. *International Review of Applied Linguistics in Language Teaching* 22(2): 115–129.

Backus, A., Doğruöz, A. S. & Heine, B. 2011. Salient stages in contact-induced grammatical change: Evidence from synchronic vs. diachronic contact situations. *Language Sciences* 33(5): 738–752. https://doi.org/10.1016/j.langsci.2011.02.004

Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B. & Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Butler, Y. G. 2002. Second language learners' theories on the use of English articles. *Studies in Second Language Acquisition* 24(3): 451–480. https://doi.org/10.1017/S0272263102003042

Cheng, L. & Sybesma, R. 2005. Classifiers in four varieties of Chinese. In *The Oxford Handbook of Comparative Syntax*, G. Cinque & R. S. Kayne (eds), 259–292. Oxford: OUP.

Cho, J. & Slabakova, R. 2014. Interpreting definiteness in a second language without articles: The case of L2 Russian. *Second Language Research* 30(2): 159–190.
https://doi.org/10.1177/0267658313509647

O'Grady, W., Kwak, H. Y., Lee, O. S. & Lee, M. 2011. An emergentist perspective on heritage language acquisition. *Studies in Second Language Acquisition* 33(2): 223–245.
https://doi.org/10.1017/S0272263110000744

Goad, H. & White, L. 2008. Prosodic structure and the representation of L2 functional morphology: A nativist approach. *Lingua* 118(4): 577–594. https://doi.org/10.1016/j.lingua.2007.01.008

Haig, G. & Schnell, S. 2010. *Annotations using GRAID (Grammatical Relations and Animacy in Discourse)*. Pdf Manual Version 7. <https://www.uni-bamberg.de/fileadmin/aspra/Publications/GRAID7.0_manual.pdf>

Heim, I. 1991. Artikel und Definitheit. In *Semantics: An International Handbook of Contemporary Research*, A. von Stechow & D. Wunderlich (eds), 487–535. Berlin: De Gruyter.

Heine, B. & Kuteva, T. 2005. *Language Contact and Grammatical Change*. Cambridge: CUP.
https://doi.org/10.1017/CBO9780511614132

Huebner, T. 1983. *A Longitudinal Analysis of the Acquisition of English*. Ann Arbor MI: Karoma.

Ionin, T., Ko, H. & Wexler, K. 2004. Article semantics in L2 acquisition: The role of specificity. *Language Acquisition* 12(1): 3–69. https://doi.org/10.1207/s15327817la1201_2

Jansen, B., Lalleman, J. & Muysken, P. 1981. The alternation hypothesis: Acquisition of Dutch word order by Turkish and Moroccan foreign workers. *Language Learning* 31(2): 315–336.
https://doi.org/10.1111/j.1467-1770.1981.tb01387.x

Jarvis, S. & Pavlenko, A. 2008. *Crosslinguistic Influence in Language and Cognition*. London: Routledge. https://doi.org/10.4324/9780203935927

Johanson, L. 2002. Contact-induced change in a code-copying framework. *Contributions to the Sociology of Language* 86: 285–314.

Kharma, N. 1981. Analysis of the errors committed by Arab university students in the use of the English definite/indefinite articles. *International Review of Applied Linguistics in Language Teaching* 19(1–4): 333–345.

Kita, S. 1995. Recommendations for data collection for gesture studies. In *Extensions of Space and Beyond: Manual for Field Elicitation for the 1995 Field Season*, D. Wilkins (ed.), 35–42. Nijmegen: Max Planck Institute for Psycholinguistics.

Lyons, C. 1999. *Definiteness*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511605789

Master, P. A. 1987. A Cross-Linguistic Interlanguage Analysis of the Acquisition of the English Article System. PhD dissertation, University of California, Los Angeles.

Mayo, M. D. P. G. & Hawkins, R. (eds). 2009. *Second Language Acquisition of Articles: Empirical Findings and Theoretical Implications* [Language Acquisition and Language Disorders 49]. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.49

Montrul, S. & Ionin, T. 2012. Dominant language transfer in Spanish heritage speakers and second language learners in the interpretation of definite articles. *The Modern Language Journal* 96(1): 70–94. https://doi.org/10.1111/j.1540-4781.2012.01278.x

Moro, F. 2016. Dynamics of Ambon Malay: Comparing Ambon and the Netherlands. PhD dissertation, Radboud University, Nijmegen.

Pires, A. & Rothman, J. 2009. Disentangling sources of incomplete acquisition: An explanation for competence divergence across heritage grammars. *International Journal of Bilingualism* 13(2): 211–238. https://doi.org/10.1177/1367006909339806

Polinsky, M. 2006. Incomplete acquisition: American Russian. *Journal of Slavic Linguistics* 14(2): 191–262.

Sánchez, L. 2004. Functional convergence in the tense, evidentiality and aspectual systems of Quechua-Spanish bilinguals. *Bilingualism: Language and Cognition* 7(2): 147–162.
https://doi.org/10.1017/S136672890400149X

Schmid, M. S. 2011. Contact x time: External factors and variability in L1 attrition. Modeling bilingualism: From structure to chaos. In *Modeling Bilingualism: From Structure to Chaos. In Honour of Kees de Bot* [Studies in Bilingualism 43], M. S. Schmid & W. Lowie (eds), 155–176. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.43.11sch

Silva-Corvalán, C. 1994. *Language Contact and Change: Spanish in Los Angeles*. Oxford: OUP.

Sloetjes, H. & Wittenburg, P. 2008. Annotation by Category: ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, 816–820. <http://www.lrec-conf.org/proceedings/lrec2008/pdf/208_paper.pdf>

R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>

Thomas, M. 1989. The acquisition of English articles by first-and second-language learners. *Applied Psycholinguistics* 10(3): 335–355. https://doi.org/10.1017/S0142716400008663

Weinreich, U. 1979. *Languages in Contact: Findings and Problems*. The Hague: Mouton.
https://doi.org/10.1515/9783110802177

Effects of first language attrition on heritage language input and ultimate attainment

Two generations of Turkish immigrants in the UK

Tuğba Karayayla

Ankara Yildirim Beyazit University

This study investigates the spoken performance of UK based adult heritage speakers (HSs) and first generation immigrants (LBs) of Turkish descent regarding the distribution of evidentiality encoded in the past tense system of Turkish. Additionally, we trace the effect of input quality and quantity on the development of the heritage language (HL). The analyses suggest that the ability to differentiate between different evidential structures is unstable in HL grammars. There are no signs of qualitatively modified input available to the HSs through the LBs. This performance is instead linked to the amount of input, which has been found to be insufficient to compensate for the detrimental effects of early bilingualism on the HL. Findings are discussed within the premises of available approaches to heritage language bilingualism.

Keywords: heritage language bilingualism, input quality and quantity, Turkish immigrants in the UK

1. Introduction

Heritage language bilingualism is a special kind of bilingualism with respect to acquisition conditions and the sociolinguistic environment of the speakers (see Montrul 2016a). In the current study, heritage speakers (HSs) are the children of the first generation of immigrants and were either born in the L2 country or immigrated with their parents at a very young age (usually before age 5). These speakers acquire the minority/immigrant language at home and achieve bilingualism depending on the timing of the L2 acquisition which can take place either from birth or with the start of (pre)schooling in the L2 country (Montrul 2016a; Rothman 2009). Unsurprisingly, these speakers ultimately become dominant in the L2, the societal and the educational language, as the L1 use remains rather limited and is

restricted to a certain number of domains (Benmamoun, Montrul & Polinsky 2013; Montrul 2008).

The development of heritage languages (HLs) in both children and adults have been studied extensively in the last decades. Although a HL might develop age-appropriately during early childhood and obtain native-like levels of L1 proficiency in adulthood (Montrul 2016a), it is not uncommon to find studies reporting less target-like grammatical behavior in childhood that lags behind the monolingual performance. This is presumably because of the shift in the language exposure patterns during school years (e.g. La Morgia 2011; Montrul 2008; Silva-Corvalán 2016). As evidence points out, this pattern might continue into adulthood. There is indeed a large body of research reporting that HSs might not fully converge on adult-like linguistic development in the HL, especially in the area of nominal and verbal morphology in both production and comprehension (e.g. Albirini, Benmamoun & Saadah 2011; Montrul 2008, 2010, 2016a; Montrul, Bhatt & Bhatia 2012; Montrul, Davidson, Fuente & Foote 2014; van Osch & Sleeman 2016). It has also been shown that the degree of this attainment might show differences across various linguistic domains or linguistic categories within the same domain (Albirini et al. 2011; Montrul 2008, 2009; Montrul et al. 2012). For example, Montrul (2009) reported maintenance of tense and aspect to a higher degree than that of mood in both oral and written production of adult Spanish HSs living in the US.

Among the successful attempts accounting for these HL divergences, incomplete acquisition (Montrul 2008) and language attrition (Polinsky 2011) explanations have prevailed in the literature. In both approaches, the linguistic categories are presupposed to be available in the input. This assumption is challenged by Rothman's (2007) input claim: the inherent properties of the input that HSs are exposed to might have changed due to attrition (and other possible reasons, such as shift, CLI and diachronic change) in the parental generation which can provide an alternative explanation to the non-convergent L1 knowledge of HSs at least for some grammatical categories (also Pascual y Cabo & Rothman 2012; Pires & Rothman 2009; Rothman 2009).

The present study primarily aims to test Rothman's (2007) qualitatively different input hypothesis by comparing the L1 performance of both adult HSs and first generation immigrants of Turkish descent in the UK to that of monolinguals. The investigation is carried out in a grammatical category known to be vulnerable in heritage bilingual populations: evidentiality (Aarssen 2001; Arslan, Bastiaanse & Felser 2015; Arslan, De Kok & Bastiaanse 2017; Karakoç 2007). This approach allows us to estimate whether HL input is of native-like quality in terms of evidential structures and evaluate further the extent of individual variation in ultimate attainment with respect to quality/quantity of input received by HSs and amount of further L1 use.

2.   Measuring input quantity and quality

Quantity of input refers to amount of the language input provided to a monolingual or a bilingual child. Acquiring two languages limits the amount of input a bilingual child can receive in each language compared to a monolingual child (Paradis & Genesee 1996; Scheele, Leseman & Mayo 2010). Despite some inconsistencies, the great majority of research conducted with bilingual/HL speakers points to the predictive role that the amount of input received in L1/L2 plays in the target language's lexical and grammatical development (Ågren, Granfeldt & Thomas 2014; De Houwer 2007; Mueller-Gathercole 2007; Hoff et al. 2012; La Morgia 2011; Unsworth et al. 2014).

Input quantity data is commonly examined through detailed parental questionnaires (e.g. Gutierrez-Clellen & Kreiter 2003; Jia & Paradis 2014; Paradis 2011). Although the basic approach in these questionnaires is very similar, i.e. asking parents a series of questions about the target language use/exposure patterns of their children, input tends to be quantified differently across studies. This lack of an agreed sound quantification method may result in inconsistent findings. Challenging traditional measures, recent research has proposed more reliable calculations for length of exposure as well as current language exposure by means of a measure called the 'Bilingual Language Exposure Calculator' (BILEC) (Unsworth 2016). BILEC considers different domains of language use in a child's environment, such as home, school and activities, and approximates the overall amount of exposure on the basis of the proportion of time spent in each domain with different input providers. All these calculations are carried out for the time the child is awake by taking the age-specific waking hours of the child into consideration. The cumulative language exposure is then calculated by adding up the language exposure values estimated over the individual years until the age at testing. Although this technique improves the traditional measures, some caution is warranted as these are still approximations based on individual reports rather than real measures and observations (Unsworth 2016).

Input quality on the other hand, refers to how rich and diverse the input is in terms of linguistic structures and vocabulary (Paradis 2011). Monolingual language acquisition has mainly associated it with the socio economic status (SES) of the family (Hoff 2006) and/or the education level of the parents. The underlying assumption is that parents and especially mothers with a higher level of education (high SES) would engage their children in activities such as reading books, more frequently, providing richer and more diverse input (Hoff 2006). This relationship between SES and input quality through home literacy activities such as storytelling, book reading, educational TV watching is, however, inconclusive in bilingual/HL contexts (Leseman, Scheel, Mayo & Messer 2009; Scheele et al. 2010). Apart from

these, how well the input provider speaks the language (Gutierrez-Clellen & Kreiter 2003; Unsworth 2013), the time spent in the L1 country (De Houwer 2009), and the number of input providers (Gollan, Starr & Ferreira 2014), have all been suggested to increase the chance of hearing more diverse and rich HL input. Reflecting these previous results, a composite variable called 'richness' has been proposed and tested in both HL (Jia & Paradis 2014) and early L2 acquisition (Paradis 2011). This variable is derived from answers given to a series of questions in the questionnaire where parents are asked whether their children participate in activities such as TV watching, book reading, typing in computer and how much of these involve L1 versus L2. This kind of engagement with the HL mainly outside home and school contexts has been shown to predict HL development in both child and adult HSs (Kondo-Brown 2005).

Much as previous research demonstrates a 'causal relationship' between input and language development, it has been argued that input should not be considered the only factor explaining the whole bilingual/HL acquisition process (Long & Rothman 2014). Factors such as the timing of acquisition (Ågren et al. 2014; Tsimpli 2014), age of onset of bilingualism (Montrul 2008), nature of the properties (e.g. transparency, complexity) (Ågren et al. 2014; Mueller-Gathercole 2007) may also play roles in bilingual language development. It has been suggested that this causal relationship might hold during the early years of development only and then fade away once children have had enough input (critical mass) to acquire the property (Aksu-Koç, Terziyan & Erguvanlı-Taylan 2014; Mueller-Gathercole 2007). In the case of HSs, however, who do not have a large linguistic community to rely on or diverse opportunities to use the language further, such examination may explain the extent of variation in the HL ultimate attainment with respect to input modifications if not the whole acquisition. As proposed by O'Grady, Lee & Lee (2011: 23) "[a] promising source of insights into heritage language learning comes from the broader study of the role of input in language acquisition".

3.   Effects of defective input on HL development

The term incomplete acquisition is used to capture the divergent L1 representational system found in HSs that might be a result of simply failing to acquire/master the L1 structures because the L1 input was quantitatively not sufficient and/or the acquisition process was interrupted by the extensive exposure to L2 at an early age (Montrul 2016a). As far as this representational system is concerned, one factor which may be of additional importance relates to qualitatively distinctive input conditions that HSs might have been exposed to (Pires & Rothman 2009; Rothman 2007, 2009), and/or HSs' sensitivity to the spoken variety around them

(Treffers-Daller, Daller, Furman & Rothman 2016). If a property is not available to HSs, such as in the case of inflected infinitives in Brazilian Portuguese which disappeared from the colloquial varieties in Brazil (due to diachronic change), and can only be acquired via formal instruction, HSs that do not have access to L1 education can only be expected not to show knowledge of that property (Pires & Rothman 2009). Alternatively, the property might be available in the input but in a qualitatively modified form due to attrition in the parental generation (Rothman 2007; Verhoeven 2004). Under such circumstances, HSs' acquisition cannot be labelled incomplete as it simply reflects the property of the language they are provided with (Pascual y Cabo & Rothman 2012), i.e. their acquisition is a complete acquisition of the language variety which is/was available to them. This is one of the reasons why some researchers express their concerns with what the term *incomplete acquisition* entails as they consider the HL development as a continuum and an approximation to the norms of the spoken variety around them, rather than something that lags behind the monolingual variety (Kupisch & Rothman 2018; Putnam & Sánchez 2013; Treffers-Daller et al. 2016). Given all these, Rothman (2009) and Pascual y Cabo and Rothman (2012) caution against evaluating HL competence divergences from an incomplete acquisition perspective before assessing the HL input conditions and suggest including a control group of first generation immigrants for future studies as an alternative account, to control for effects of altered input.

If age of onset of bilingualism (AoA) is a strong predictor of L1 attrition (Bylund 2009; Montrul 2008), the L1 performance of late bilinguals (LBs) and HSs should, in principle, differ primarily due to their AoA. As argued by Kaltsa, Tsimpli and Rothman (2015), any similarity found in the performance of these groups that differs from that of monolinguals can be attributed to the effect of "attrited" input or convergence on the spoken variety around them (Treffers-Daller et al. 2016). This has already been evidenced in some relatively recent studies. Pascual y Cabo (2013) for instance, showed that one of the reasons why Cuban HSs in the US allow an ungrammatical optional use of agentive syntax with *gustar*-like verbs (class III psych-predicates) was because of the loss of a property (dative marking) in the L1 of the input providers. This property is relevant for the production of this class of verbs, corroborating Rothman's (2007) input claim.

Montrul and Sanchez-Walker (2013) compared the oral performance of different bilingual groups in the US: first generation Mexican immigrants, child and adult HSs, and monolinguals in Mexico. They found significant omission rates of the property (differential object marking, where English lacks a counterpart) in the production of child and adult HSs as well as LBs at group level suggesting vulnerability of the property to incomplete acquisition and attrition. To further address the incomplete attainment, they divided the HL groups into two (omitters versus non-omitters) based on their accuracy performance. A higher level of L1 use was

associated with better performance. As some attrition was found in the first generation as well, they speculated that variable performance among adult HSs who use their L1 with their parents and older relatives could be related to input quality they received from attrited parents. Similar conclusions were found by Montrul (2016b) in an investigation of overt pronoun use which uses a similar design of participant groups. As Montrul and Sanchez-Walker (2013) claim, these findings demonstrate a complex relationship between background factors and their contribution to ultimate HL attainment, therefore signifying the roles input quantity/quality play in addition to the effects of other factors such as CLI (see Kaltsa et al. 2015 for a similar argumentation).

In summary, as stated by Sorace (2014), despite extensive references in the literature to qualitatively different input that HSs might be exposed to, it is quite surprising that not many studies examined this. With this line of thinking, we hope to provide new empirical data and supplementary findings to the existing breadth of literature reviewed above by investigating HL ultimate attainment from all these aspects – defective input, incomplete acquisition and attrition – in a non-Indo-European HL (Turkish) in contact with English. More precisely, we do not only control for the qualitatively different input available to HSs, but also trace the effect of input reductions over time on the development of the HL. This enables the best identification of the sources of non-convergent HL performance.

4.   Evidentiality in Turkish

Evidentiality is a grammatical indication of how the information is acquired, i.e. it refers to the source of knowledge in a proposition (Aikhenvald 2004; DeLancey 2001; Johanson 2006; Lazard 2001; Schroeder 2000). Although there are ways to refer to sources of knowledge in all languages, evidentiality can be considered as grammaticalized in a language only if the grammatical system includes markers which semantically or pragmatically refer to the source of knowledge (Lazard 2001). While Turkish grammaticalizes evidentiality in its complex tense-aspect-mood (TAM) system with verbal morphology, English does not (Aksu-Koç 1988, 2000, 2009; Göksel & Kerslake 2005).

In references to past events, a native speaker of Turkish has to choose between two different verbal suffixes: the direct experience evidential (Dexp) form -*DI* and the indirect experience evidential (INDexp) form -*mIş*[1] which additionally

---

1.   Previous literature proposes an additional separate marker represented as -*(I)mIş* (Csato 2000; Johanson 2006). This marker works as a copula marker or as a clitic attaching to nominal predicates and/or already inflected verbs to form complex verbs (Sezer 2001). This form is claimed

subsumes inferential (infE) and reportative evidential (repE) usages. Uttering *Bina yan-dı* ("The building burnt down") with the Dexp marker *-DI* would mean that the speaker is a witness of the event, and thus has access to this information through the source knowledge of direct perception (visual access) and/or participation. Uttering *Bina yan-mış* ("Apparently/I have been told that the building burnt down") with the INDexp marker *-mIş* on the other hand, is an indication that the speaker did not witness the event directly but acquired the information via one of two sources: inference or reportative (hearsay). In the former, the speaker does not see the building burn down (nonwitnessed) but has access to physical or visual evidence/resultatives, i.e. remains after the fire, allowing the inference or the logical reasoning of the event (Aksu-Koç 1988; Slobin & Aksu 1982). In the latter, on the other hand, this information is acquired via linguistic reports and/or third parties, and again this is a nonwitnessed event uttered/reported based on what was heard/read. Both forms of *-mIş* are considered as representing "information new for unprepared minds" (Slobin & Aksu 1982: 196–198) and "can be read as 'on the basis of what I have just found out, I assert that…'." (Aksu-Koç 2000: 18). The Dexp is rather considered as encoding "a neutral perspective" which "presents the event as representation of reality" (Aksu-Koç 2009: 533). Both forms of -mIş can be distinguished from the Dexp by the modal meaning of not witnessing the event (Aksu-Koç 1988). InfE forms require the tense to be past (anterior) with the effects of the event, i.e. remains after the fire, visible at the time of observation (resultative), as in the example above (Aksu-Koç 1988; Bacanlı 2008; Johanson 2006; Şener 2011). The repE form, on the other hand, behaves similar to the Dexp in terms of its temporal/aspectual function: they both refer to anteriority and are compatible with both specific, e.g. *yesterday*, or non-specific, e.g. *recently*, time reference adverbs (Bacanlı 2008; Şener 2011). This means that they can both mark definite past or present perfect (see examples in Section 8).

5.   Evidentials in monolingual and bilingual contexts

Longitudinal investigation of evidentiality pioneered by Aksu-Koç (1988) indicates the following order in monolingual acquisition: Dexp (around 1;6–2;0) < infE (2;0–2;6) < repE (2;0–3;0). Complementary data provided by Aksu-Koç (2000) indicate that the first appearance of *-DI* is a reference to verbal change of states, as well as

---

to be a pure evidential marker conveying hearsay which does not necessarily mark aspect or tense unless the time reference is specified by the discourse context and/or time adverbials such as *dün* 'yesterday' (Göksel & Kerslake 2005; Sezer 2001). In the current research the focus will only be on the verbal past tense suffixes which mark both tense-aspect and evidentiality.

completed actions, which then extends to descriptions in the remote past. As the data showed, the suffix -*mIş* on the other hand, first appears in nominal predicates as new information to mark the current states of the entities and includes a surprise element on the basis of direct perception. This extends gradually to picture stories, imitation of adult story-telling, pre-tense references to "physical or emotional states of third parties" as well as imaginary role-plays (Aksu-Koç 2000: 21). Around the age of three, children use these forms correctly to refer to past events (Aksu-Koç 2009). However, the differentiation between witnessing vs. nonwitnessing modality on the basis of different sources of information, i.e. evidentiality, does not appear before the age of 3;6. Children still show divergences in these usages, especially in the production of the INDexp until ages of 6–7 years (Aksu-Koç, Ogel-Balaban & Alp 2009; Ozturk & Papafragou 2008a, 2008b, 2016). There is a similar asymmetry in the comprehension of the two perspectives (direct versus indirect) (Aksu-Koç 1988; Ozturk & Papafragou 2008b, 2016). The correct production of evidential morphemes, however, precede their overall comprehension as evident in tasks where 5–7 year-olds failed to attribute an utterance to the correct speaker who had access to that information (Ozturk & Papafragou 2008a, 2016).

This has been shown to possibly relate to the theory of mind (ToM), acquisition of knowledge, or perspective-taking abilities developing independent from language in a series of studies in Turkish (Aksu-Koç 2009; Aksu-Koç et al. 2009; Ünal & Papafragou 2016). That is to say, children's assessment of the information source available to others might not yet be determined by the linguistic encoding of source in their languages although they produce these forms (Aksu-Koç 2009). There is, however, evidence that shows that the ability to assess their own non-linguistic source of information precedes that of production of evidential morphemes (Ozturk & Papafragou 2008a, 2016; Ünal & Papafragou 2013, 2016).

The fact that the direct experience perspective/production develops earlier presumably relates to the transparency of the forms in the input, and the mapping of these forms onto conceptual source functions (after information source concepts develop) (Ozturk & Papafragou 2016; Ünal & Papafragou 2013). In terms of source marking, while -*DI* is a unifunctional marker, -*mIş* has more than one function. As proposed by Slobin (2001), forms that map into more than one meaning might be more difficult to acquire. Although evidential -*mIş* is not as transparent as -*DI* in the input, a longitudinal study conducted by Aksu-Koç et al. (2014) has indicated that both of these forms are acquired from the input. Here, the authors examined the distribution and acquisition of TAM suffixes (both modal and non-modal meanings) in two Turkish mothers' and their children's utterances between ages 1;3–2;6. They indicated that input provides rich and diverse structures specified for form-function mappings. For less transparent structures, the interaction of frequency with transparency becomes more important, and children acquire these

multifunctional structures only gradually (Aksu-Koç et al. 2014; see also Ünal & Papafragou 2013).

Although limited in number, evidentials have received attention in bilingual contexts as well. Earlier research was concerned with whether young HSs would converge on monolingual narrative structure in terms of a consistent temporal verb choice and discourse functions. Theoretically, it is possible to retell a story in all three modalities: *-mIş*, *-DI* and the present *-(I)yor*. Some forms might be more appropriate and traditional depending on the reference taken, i.e. the picture book of the story, the film of the story seen before, or the read-aloud story (Karakoç 2007). As noted by Aarssen (2001: 213) "[o]ne way of organizing a narrative is to maintain an anchoring tense throughout the text" and use it appropriately in the required discourse.

Taking these as the main criteria, Aarssen's study conducted with 140 Turkish HSs aged 4–10 in the Netherlands showed that while younger HSs presented a high percentage of "unmotivated" (serving an inappropriate discourse function) and inconsistent (not sticking to one anchor form) tense shifts in their narratives of frog stories, the rate of these shifts decreased with an increasing age. The use of both forms of past tense appeared only after age six with a main preference of present *-(I)yor* as the anchored tense. Only ten year-olds used the past tense forms as the anchor tense in their narratives but still with some unmotivated shifts. Unlike these findings, however, none of the Turkish-German bilingual subjects (aged 5–8) in Karakoç's (2007) investigation took the repE as their basis (anchor) to retell the read-aloud story "Snow White". This form occurred only in unmotivated shifts. This kind of inappropriate contextual use seems to persist into adult HS performance. Arslan et al. (see below) have recently shown that evidentials are subject to attrition/incomplete acquisition in adult immigrant groups due to their inherently complex components. An investigation of the spoken performance of adolescent HSs (aged 16–18) in the Netherlands, for instance, reported a tendency to use the direct evidential inappropriately in obligatory indirect evidential contexts, an indication of a failure to differentiate between different sources of information (Arslan & Bastiaanse 2014, cited in Arslan, Bastiaanse & Felser 2015). This led the authors to two conclusions: the HSs might have lost the evidential meaning of the Dexps and used them only to refer to past events and that they retained the evidential meaning of the INDexps, at least to some limited degree, as these forms did not appear in place of the Dexps. As shown in a later study, evidentials in the performance of adult HSs were not only affected in production but also in comprehension. This is demonstrated by reduced sensitivity to evidential violations in comparison to time-reference violations and to monolingual performance (Arslan, De Kok & Bastiaanse 2017). Interestingly, this reduced sensitivity did not differ between direct versus indirect evidentials.

Finally, in an eye-tracking experiment, Arslan, Bastiaanse & Felser (2015) compared online processing of evidentials in both Turkish-German LBs and HSs with respect to AoA. As their findings showed, both LBs and HSs responded to direct evidentials less accurately and more slowly in comparison to monolinguals. For indirect evidentials, surprisingly, no significant differences were revealed between groups (including monolinguals). The authors evaluated this unexpected result in terms of providing support to their previous conclusions: the fact that Dexps were affected while INDexps remained intact seems to indicate that both groups of bilinguals lost their sensitivity to the evidential value of the Dexp but retained that of the INDexp.

In short, evidential forms seem to be affected in both child and adult HSs. Potential sources of this L1-divergent performance have, however, remained rather unclear. Despite not being in the scope of the study, Arslan, Bastiaanse & Felser's (2015) findings appear promising to show the influence of qualitatively different input due to reported attrition in the L1 of the LBs. This is yet to be identified.

6.   Research questions and hypotheses

The current research aims to integrate heritage language acquisition/ultimate attainment and first language attrition studies to address the following questions:

1.   Does the input available to the Turkish-English HSs in the UK show any qualitative differences from monolingual input in terms of evidential structures?
2.   Do the adult HSs differ in their overall accuracy of evidentials from that of Turkish monolingual speakers living in Turkey and from late bilinguals (LBs) in the UK?
3.   How do the quantity/quality of the input which HSs were exposed to at different stages of their acquisitional process, as well as their language use in adulthood, relate to their overall accuracy in the use of the evidentials?

As the direct evidential is the default past tense form and shares a surface similarity with the English past tense marker, it is expected to be more resistant to selectivity in immigrant groups when compared against indirect evidentials that require special pragmatic and semantic contexts. If this triggers any sort of attrition/CLI in the LBs, based on Rothman's (2007) input claim this should be reflected in the HSs' performance. We therefore assume that qualitatively modified input received by the HSs will help explain HL divergences. Additionally, based on previous literature, being multifunctional and less transparent, indirect evidentials would be more difficult and require more input to be acquired. They are therefore expected to be more vulnerable to input effects and be affected to a greater degree in comparison with direct evidentials in the HS performance.

7.   Participants

The spoken performance of 31 UK-born adult HSs, 31 first generation immigrants (LBs) representative of the parental generation of the HSs and 44 monolinguals as a control group (CG) in Turkey was investigated. Table 1 and 2 below provide basic background information about the participants.

Table 1. Basic background information

	Groups	HS	LB	CG
age	group size (N)	31	31	44
	mean	23.35	41.06	33.81
	range	18–43	25–65	18–66
	SD	5.88	8.01	11.8
AoA	mean	2.8	22.35	N/A
	range	0–5	12–12	N/A
	SD	1.07	7.09	N/A
LoR	mean	23.35	18.7	N/A
	range	18–43	8–40	N/A
	SD	5.88	7.44	N/A
Gender	female (N)	17	16	22
	%	54.84	51.61	50
	male (N)	14	15	22
	%	45.16	48.39	50

Due to the sociological diversity among the Turkish-speaking community in the UK, participants were selected with care to control effects of any other known native language such as Kurdish or Arabic. The regional variation in Turkish, mostly limited to phonology and lexis (Lytra 2012) would not affect the knowledge of evidentiality. Nevertheless, it has been controlled as much as possible by including a CG representative of the experimental group participants in terms of their city of birth in addition to gender, age and educational background.

Table 2. Educational background of the participants

	Total	University		High school		Secondary school	
		N	%	N	%	N	%
HS	31	25	74.19%	8	25.81%	0	0.00%
LB	31	17	54.84%	13	41.94%	1	3.23%
CG	44	25	56.82%	16	36.36%	3	6.82%

Three HSs were simultaneous bilinguals mainly due to the effect of TV and older siblings. Two participants learnt English upon starting primary school at the age of 5. The rest became bilingual at around age 3 once they started preschool ($M = 2.8$). They were all exposed to Turkish from birth.

The LBs were selected mainly on the basis of their age of arrival in the host country (AoA > 12, $M = 22.35$), to ensure full acquisition of L1 in a monolingual environment prior to emigration, and on their length of residence (LOR > 8, $M = 18.7$) (Schmid 2011). The current age was limited to 65 to prevent advancing age effects on language abilities.

8. Instruments

8.1 Sociolinguistic questionnaire

(Socio)linguistic information about the participants was collected via two versions of a SQ adapted from Yılmaz (2013) which was developed based on the test battery proposed by Schmid.[2] The language input/output questions in the HS version were developed based on the extensive parental questionnaires in Jia and Paradis (2014) and BILEC (Unsworth 2016). Both versions of the questionnaires are available in Turkish at languageattrition.org.

8.2 Semi-structured interview (INT)

Evidential forms were captured naturally by creating special evidential contexts that required participants to tell stories related to their past experiences. The interview consisted of seven questions, the first three of which were designed as warm-up questions and the next four to elicit evidential forms. To elicit Dexp, the participants were asked to tell a story they experienced/witnessed in the recent past. For repE, they were asked to talk about specific childhood stories that they do not directly recall but were told of by their parents. They were additionally asked to tell a story they did not witness but heard recently in the news or from their acquaintances/friends. As exemplified from the data below, stories the participants told as a response to these questions might refer to a non-specific time as in (1a) and (2a) or a specific time as in (1b) and (2b). Note that in all these contexts, the verbal suffixes *-DI* and *-mIş* mark the described event as something happened in the past and the information source is either visual evidence/participation (1a) and (1b) or verbal reports (2a) and (2b).

---

2. See languageattrition.org.

(1) a. *Bir defa uçak-ta   başüstü dolap-ları   aç-ıl-dı.*
         once   plane.LOC overhead cabinet.3PL.POSS open.PASS-D.PASt
         "Once on the plane the overhead cabinets were opened."
     b. *Dün   Türkiye bilet-ler-imiz-i   al-dı-k.*
         yesterday Turkey   ticket.PL-1PL.POSS-ACC buy.D.PAST-1PL
         "Yesterday, we purchased our tickets to Turkey."

(2) a. *Bir kere yatak-tan düş-müş-üm.*
         once   bed.ABL   fall.M.PAST-1SG
         "They say, I fell off the bed once."
     b. *Dün   sınav-dan sonra çok ağla-mış   kız-ım.*
         yesterday test.ABL   after a lot cry.M.PAST-3SG daughter.1SG.POSS
         "Apparently, my daughter cried a lot after the test yesterday."

InfE forms were elicited as a response to a hypothetical theft scenario. In the scenario a thief broke into the house and left mess everywhere when the participants were on vacation. Their job was to call and give details to the police (the investigator) about the situation in the house. This question is expected to elicit infE forms as there is no witnessing of the event by the participant. It can only be inferred from resultative states, such as a broken window, relocated sofas and so on, as in (3).

(3) *Cam-ı   kır-mış,   dolap-lar-ı   karıştır-mış,*
     Window.ACC break.M.PAST, wardrobe.PL-ACC disorganize.M.PAST,
     koltuk-lar-ı çek-miş.
     sofa.PL-ACC pull.M.PAST
     "Apparently, s/he broke the window, disorganized the wardrobe and relocated the sofas."

8.3   Picture description task (PD)

This task was designed to elicit infE forms only. Participants were shown five real pictures of incidents/events in colored A4 size form collected from online versions of the Turkish newspapers *Milliyet* and *Hürriyet.* All the pictures show resultative states of what happened, allowing the inference of the event. Picture one (Gaziantep'te sel felaketi 2014) depicts a flooded village where people are trying to collect remains after the flood that harmed their houses and animals. Picture 2 (Çoban 2014) shows a crane which rolled over a building and damaged different parts of the building. In Picture 3 (Yürürken üzerine duvar yıkıldı, 2014), there is an elderly lady wounded, and she is being offered first aid and emergency care. Picture 4 (Tafolar 2014) shows cutting down of a big, old tree on a busy street and its removal by a team of workers. Finally, Picture 5 (Meriç suyunun hapsettiği mahalle:

Karaağaç, 2014) shows some physical damage to vehicles and a bridge over a river presumably caused by excessive rainfall and rise in the water level. Each participant was asked to describe them as in (4), and their performance was audio-taped.

(4) *Bu   resim-de   sel   olmuş,*   *insan-lar-ın   hayvan-ları*
    this picture.LOC flood occur.M.PAST, person.PL-GEN animal.3PL.POSS
    öl-müş.
    die.M.PAST
    "In this picture it seems that a flood occurred and those people's animals died."

9.   Transcription and coding

The data consists of 155471 words (33.63 hours). Transcription of the data (INT+PD) was done according to CHAT conventions (MacWhinney 2000) using conventional orthography and spelling. No task differentiation was made in data coding as the PD was used as a supplementary task to the INT to collect infE forms only. Individual recordings lasted 10–35 minutes (*M* = 19.04).

The data was coded for Tense-Aspect-Modality (TAM) by using the simplified version of the framework[3] used in Aksu-Koç (1994: 339) shown below in order to investigate evidentiality encoded in the past tense system. Context-appropriate usages and substitution errors were of interest in this respect.

Table 3. TAM data coding framework

Verb form	
present tense	existentials (substantives)
	present -*(I)yor*
	habitual -*(A)r/(I)r*
dpast tense	verbal suffix -*DI*, and auxilary *ol*- for nominal predicates)
mpast tense	verbal suffix -*mIş*, and auxilary *ol*- for nominal predicates)
future tense	future *(-AcAk)*

---

3.   Note that the coding table was used to achieve consistency in the coding of the TAM markers relevant for the current analysis only. It is thus not exhaustive and does not reflect the entire TAM system of the Turkish language. For exhaustive lists and explanations see Göksel & Kerslake (2005) or Sezer (2001).

Following this schema, each inflected verb[4] (total $n = 20181$) was coded for the main anchored tense (e.g. dpast) according to the suffix (e.g. *-DI*) used to inflect the verb, and for its evidentiality status (zero versus direct experience/indirect experience). For the purposes of the current study, all *non*-past usages were coded as 'zero' in terms of their evidentiality status. For past[5] usages, evidentiality was either coded as *eviddi* ($n = 5424$) or *evidmis* ($n = 3224$). *Evidmis*[6] was further coded as inferential[7] ($n = 2036$) or reportative ($n = 1188$) depending on the context created in the interview questions and picture description task. Each context-appropriate evidential used was coded as accuracy = 0. Replacement errors, indirect evidentials substituted with direct evidentials, were coded as accuracy = 1. If it was reversed, this was coded as accuracy = 2. There was only 1 (0.01%) case of the latter, as opposed to 197 (6.11%) cases of the former. There were no omissions or substitutions with other markers. The data file with the errors coded as described above was then used to carry out statistical analyses on the accuracy performance of the groups in R statistical platform (see Results).

---

4. Note that auxiliary *ol-* ('to be, to become') "displays the formal structure of the inflection of complete verbs with the Tense suffixes it hosts" (Sezer 2001: 15). Given this, nominal entities followed by *ol-* as in *ameliyat olmuş* ('apparently/they say, she had an operation') have been treated as complete verb forms and coded for tense and evidentiality status.

5. As the main focus of the indirect evidential investigation was reportative and inferential contexts in references to past events, other contexts that the indirect evidential marker *-mIş* creates, such as assumption and fairy tale contexts, counterfactual contexts (Csato 2000) and other usages of the same marker without evidential readings included in Bacanlı (2006, 2008) and Johanson (2000), were not counted. Similarly, well-recognized and common historical events described by direct evidential forms despite being indirect and not referring to any first-hand experience (Johanson 2006; Schroeder 2000) or present meanings of the same form when used with psychological verbs (see Sezer 2001: 10 for details) were also not included in direct evidential counts. All these usages were coded as "other markers" and excluded from the counts. There were 41 cases of the former (*-mIş*) as opposed to 235 cases of the latter (*-DI*).

6. The information source marked by the INDexp marker is proposed to be "limited to main clauses with a stated, contradictable content" and thus does not extend to subordinate clauses in Turkish (Johanson 2006: 81). There is, however, meaningful evidence showing that the INDexp marks the information source in certain types of finite clauses, such as bare finite subordinate clauses, finite clauses constructed with *ki* and *diye* (Coşkun 2010), and bare subordinate clauses of the verb *de-* (Şener 2011). Following these studies, the INDexp form that marks the information source in such clauses was also considered as *evidmis* and further labelled as inferential or reportative on the basis of the context.

7. The indirect evidential contexts that have been shown to have inferential reading when used together with the auxiliary *ol-* in necessity modality *-malı* (Şener 2011) were considered as inferential.

10. Predictive variables

The following predictive input/output variables were calculated from the SQs.

10.1   Current L1 contact

Following Schmid and Dusseldorp (2010), a principal component analysis (PCA) with varimax (25) rotation was conducted on the items that relate to frequency of language use in the sociolinguistic questionnaire. Three new L1 contact variables below (Table 4) were established whose composite scores were computed for each participant as the means of the variables[8] included in each component. Internal consistency was established conducting a reliability analysis (Cronbach alpha).

Table 4. L1 contact variables

		Interactive L1 use	L1 passive exposure	L1 use with friends
heritage speakers	mean	0.77	0.55	0.43
	range	0.42–1.00	0.00–1.00	0.13–1.00
	SD	0.14	0.28	0.14
late bilinguals	mean	0.92	0.59	0.66
	range	0.75–1.00	0.17–1.00	0.18–0.95
	SD	0.08	0.22	0.2
reliability	alpha	0.789	0.649	0.603

10.2   Input quantity

This study made use of some of the calculations of BILEC (see Unsworth et al. 2014 and Unsworth 2016 for details about calculations). In order to account for domain-specific (e.g. home or school) language exposure for individual years, BILEC's calculations include the proportion of time spent in each domain by taking the age-specific waking hours of children into consideration. This is to assess the exact time spent with input providers during the time a child is awake in each domain and calculate how much language exposure the child receives during this time. These domain-specific language exposure measures are then incorporated

---

8.   The choice was made on high loadings of the items on one component. The individual variables that each composite variable included are provided in parenthesis below: (i) Interactive L1 use (L1 use with children, siblings, parents, grandparents in Turkey, other relatives in the UK, L1 use while writing to relatives in the UK and in Turkey); (ii) L1 passive exposure (non-interactive L1 use/exposure through TV, radio and music); (iii) L1 use outside home (cultural preferences for friends and L1 use with friends and neighbours).

into one exposure variable. The cumulative amount of exposure is the sum of the language exposure the child has had over the individual years.

We adapted[9] this approach to account for *past L1 experience* of the HSs between the ages of 0 and 18. This allowed us to achieve relatively reliable measures. This age range was divided into four periods corresponding to UK educational periods. These calculations resulted in four exposure and four use variables.

10.3   L1 and L2 richness (input quality)

Adapting Jia and Paradis' (2014) calculations,[10] a variable called *richness* was derived from the "activities" section of the questionnaire. Please note that despite some quantification involved in its calculations, this variable has been proposed as being more quality-oriented (Paradis 2011) due to referring to the diversity of the linguistic environment in terms of L1/L2. Calculations resulted in one L1 and one L2 richness score for each one of the four age periods per participant.

In later stages, due to highly significant correlations (Pearson) between variables for the ages 0–3 and 3–5, composite variables were created for the ages 0–5 for *exposure, use* and *L1/L2 richness*. Following this, significant correlations between each age-specific exposure and use variables resulted in three age-specific

---

9.   Based on participants' answers in the sociolinguistic questionnaire, for each linguistic domain ((1) home, (2) school and (3) outside home and school, i.e. extracurricular activities), two kinds of domain-specific variables were derived in percentages for each age period: L1 exposure and L1 use. Average approximations of waking hours for each age period were achieved based on reported findings of medical articles on sleep durations of children growing up in Europe (Iglowstein, Jenni, Molinari & Largo 2003; Mindell, Sadeh, Wiegand, How & Goh 2010; Olds, Blunden, Petkov & Forchino 2010). This way, we were able to calculate the proportion of time spent in each domain. This allowed us to incorporate the domain specific exposure and use percentages mentioned above into one *exposure* and one *use* variable for each age category.

10.   In the first stage of calculating this variable, the reported amount of time by the participants for the activities was multiplied by the reported L1 versus L2 percentage involved in it. This would give us the total number of hours of Turkish/English involved in that activity for each individual. The separate L1 and L2 richness score for each participant and each activity category was determined according to the largest value derived as a result of this multiplication. This value simply reflects the biggest number of hours spent doing that particular activity in Turkish and English respectively. All the other scores were divided by this value. This method of data normalization allowed us to evaluate each participant according to a common base. As a result, each participant received a richness score out of 1 in each category for each age period. Since there are five categories in the "activities" section, the scores each participant received from each category out of 1 was added up. This means that the highest possible L1 and L2 richness score is 5. Note that the richness scores were calculated separately for L1 and L2 and thus an L1 richness score of 4 does not mean that the score for L2 richness equals to 1.

compound *exposure + use* variables calculated as the mean value. The new compound variables are called *L1 experience* as shown in Table 5.

Table 5. L1 experience, L1/L2 richness variables

Age category		Mean	Range	SD
	L1 experience	0.72	0.25–0.96	0.18
0–5	L1 richness	1.2	0.00–2.84	0.61
	L2 richness	0.97	0.1–2.89	0.73
	L1 experience	0.39	0.16–0.65	0.13
5–11	L1 richness	0.83	0.00–2.02	0.55
	L2 richness	1.68	0.6–3.47	0.61
	L1 experience	0.34	0.04–0.57	0.15
11–18	L1 richness	1.02	0.00–2.82	0.7
	L2 richness	2.01	0.63–1.00	0.82

These values are assumed to be the most reliable in the absence of real observations and longitudinal studies.

11. Results

Speakers in the CG produced 1616 words on average, while the LBs produced 1441 and the HSs 1278 words. All verb counts were calculated per 1000 words per individual (see Schmid 2011). Past usages were coded as either dpast and mpast, while all finite non-past verb forms except present *-(I)yor* were classified as "other". Present *-(I)yor* was included separately because it is an alternative form in Turkish narratives (narrative present) to dpast and/or mpast (Aksu-Koç 1994; Karakoç 2007). This was to reflect the possibility that some participants might have told their stories using the narrative present, which might lead to dpast and mpast distribution inequalities across the groups.

As Figure 1 shows, the HSs produced a slightly bigger number of finite verbs than the other two groups. A One-way ANOVA and Tukey tests confirmed that this difference between groups was not significant ($F(2, 103) = 1.176$, $p = 0.313$). The groups differ significantly in the amount of infE ($F(2, 103) = 4.907$, $p = 0.009$) and the difference in the amount of repE[11] approached significance ($F(2, 97) = 2.827$, $p = 0.064$) respectively. No significant differences were revealed for the dpast and non-past usages (all ps > 0.05).

---

11. Four LBs and two HSs did not use any reportative *-mIş* forms. All relevant calculations concerning repE forms (e.g. accuracy) were carried out for 27 LBs and 29 HSs.

The HSs produced more infE than both the LBs ($p = 0.028$) and the CG ($p = 0.014$). Surprisingly, both the HSs and the CG used repE forms more frequently than the LBs but only the difference between the LBs and the CG approached significance (Tukey $p = 0.072$). There were no other significant differences (ps > 0.05).
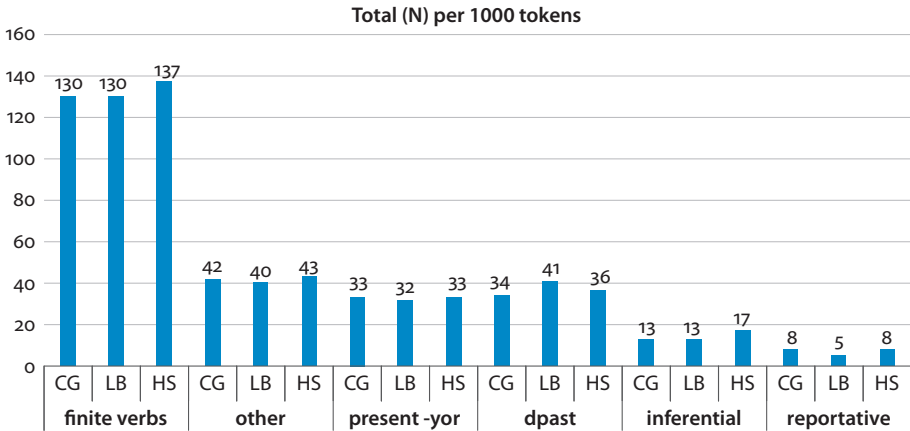

Figure 1. Mean distribution of the tenses across the groups (in N)

11.1   Evidential accuracy

Substitution errors were calculated for the evidentials only. The groups showed a tendency to substitute the direct evidential in contexts requiring the indirect evidential. In (5) for example, the participant talks about a conversation between her dad and her sister that she did not witness herself but heard from her sister on the phone. According to this story, the sister arrives home drunk. Her father is not happy with this situation and tells her they are going to talk about this in the morning. Interestingly, the participant reported the whole story in the direct evidential form as if she witnessed the event. Table 6 displays some descriptive information on the evidentiality performance of the groups.

(5) *Ev-in   iç-i-ne   gir-ince tabi ki   baba-m*
     house.GEN inside.3SG.POSS-DAT enter-CV of   course father.1SG.POSS
     hemen   anla-DI,   de-DI   "yarın   konuş-ur-uz".
     immediately understand.D.PAST*, say.D.PAST* tomorrow talk.AOR-1PL
     "As soon as she entered the house, he immediately understood and said 'we will talk tomorrow'."

Statistical group comparisons on evidential accuracy were made by calculating a number of Generalized Linear Mixed Effects Regression Models (GLMM) with the *lme4* package (Bates, Mächler, Bolker & Walker 2015) for R version 3.2.4 (R Core

Team 2016). The main motivation behind this choice over traditional ANOVAs was the fact that mixed effects models are very powerful in dealing with unbalanced corpus data (Gries 2015). What is meant by unbalanced data in the context of the current study is that some errors made might come from a limited number of sources (participants) in the entire group, or there might not be an equal number of observations for each evidentiality type per each participant (e.g. inferential versus reportative). Mixed effects models are suitable in these situations and take the variability caused by participant characteristics described above into consideration as well (Baayen, Davidson & Bates 2008). For this reason, the variable *participant* is included as a random intercept in the analysis. The model took the accuracy performance of the CG and the performance in the Dexp as the baseline.

We included *group* (three levels: CG, LB, and HS) and *evid type* (three levels: Dexp, infE and repE) as fixed effects with the *evidential accuracy* (two levels: accurate versus inaccurate) as the dependent variable. The estimates of the model showed that in general, both the group ($\beta = -1.895$, $SE = 0.27$, $z = -6.82$, $p < 0.001$) and the evid type ($\beta = -3.387$, $SE = 0.23$, $z = -14.49$, $p < 0.001$) were significant predictors. As revealed by post hoc Tukey tests, the HSs were significantly less accurate than both the CG ($\beta = -4.79$, $SE = 0.71$, $z = -6.69$, $p < 0.001$) and the LBs ($\beta = -3.36$, $SE = 0.58$, $z = -5.75$, $p < 0.001$) but there were no differences in the overall accuracy performance between the LBs and the CG (Tukey $p = 0.146$). The general performance in both infE ($\beta = -4.556$, $SE = 1.02$, $z = -4.46$, $p < 0.001$) and the repE ($\beta = -7.928$, $SE = 1.02$, $z = -7.632$, $p < 0.001$) were less target-like than in Dexp and it was less target-like in repE ($\beta = -3.272$, $SE = 0.27$, $z = -12.02$, $p < 0.001$) than in infE. In summary, the evidential accuracy performance displayed a rank order as follows: repE<infE<Dexp.

These analyses were followed by a number of similar models built for direct and indirect evidentials separately across groups. The ceiling performance in Dexp did not make it possible to carry out a group comparison but the HSs were significantly less accurate than the CG ($\beta = -4.896$, $SE = 0.77$, $z = -6.32$, $p < 0.001$) and the LBs ($\beta = -3.272$, $SE = 0.64$, $z = -5.05$, $p < 0.001$) in indirect evidential contexts. Both the CG and the LBs were equally accurate in using these forms (Tukey $p = 0.12$).

Table 6. Evidential accuracy

Group	Direct experience			Inferential			Reportalive		
	Total target mean (M)	Total replacement with indirect (M)	Accuracy %	Total target mean (M)	Total replacement with indirect (M)	Accuracy %	Total target mean (M)	Total replacement with indirect (M)	Accuracy %
CG	54.70	0.00	100.00	19.45	0.00	100.00	12.36	2.00	99.22
LB	57.84	0.00	100.00	17.81	0.00	100.00	7.64	2.40	95.39
HS	39.55	0.03	99.96	20.26	1.64	95.04	13.87	6.17	69.40

Given these results, the input available to the HSs seems to be native-like. The individual variability in the CG and the LBs performance below confirms this finding. Given that Dexp and infE performance was at ceiling in both groups, Figure 2 below shows individual variation only for the repE.

Only two out of 27 (see Footnote 11) (7.04%) LBs were outside the range of the accuracy achieved by the CG with accuracy percentages of 63.63, and 44.44. For the rest (92.96%) the accuracy ranged from 80.00 to 100.00%. The CG accuracy range was also 80.00–100.00%. The individual variability in the HSs' performance was greater. As Figure 3 demonstrates, the performance in the Dexp was at the ceiling. 17 (54.84%) HSs performed at the ceiling (in the control range) in the infE. The accuracy of the rest (45.16%) ranged from 60 to 96.55%. The least accurate two participants were 60% and 73.33% target-like.
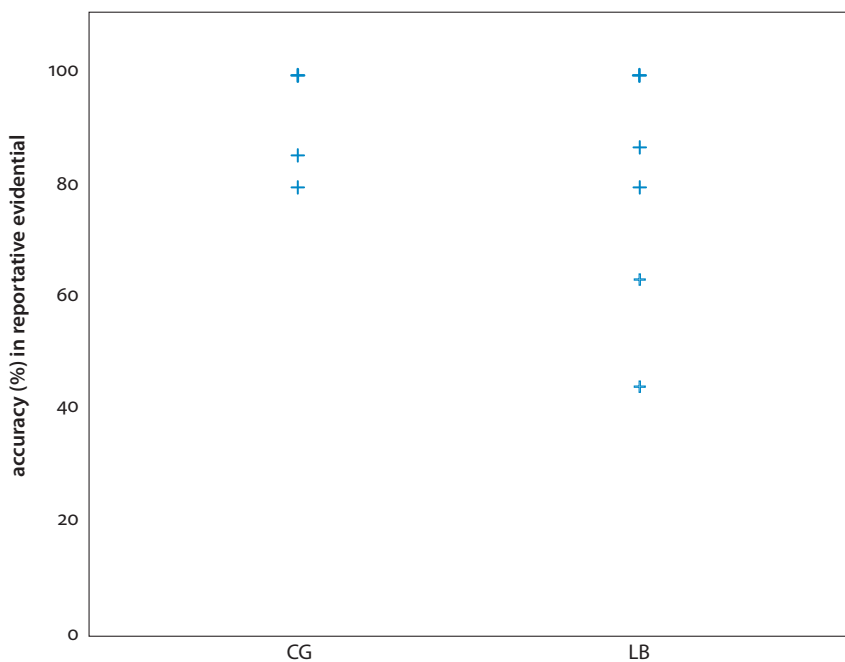
Figure 2. Individual variation in repE accuracy

For the repE on the other hand, 100% target-like performance was achieved by only seven out of 29 participants (24.13%). For the rest (75.87%), the performance ranged from 6.66% to 94.11%. Overall, 14 (48.27%) participants remained below the control range in the repE. The four participants who were at the bottom had accuracy percentages of 6.66, 10.00, 10.52 and 30.00.

Nine participants (29.03%) performed within the control range in the indirect evidentials overall.
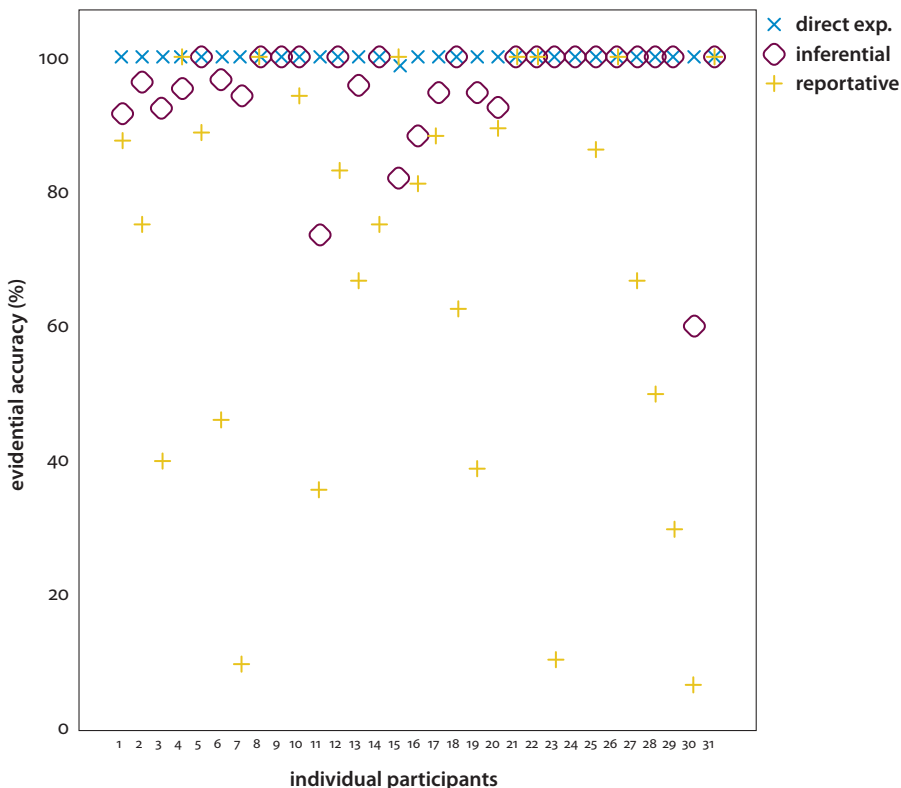
Figure 3. HS individual variation

In sum, the HSs' performance in the indirect evidentials clearly diverges from that of the monolinguals. This variability cannot be explained by any qualitatively modified input conditions. Nonetheless, differences in input quantity, L1/L2 richness and current L1 contact might prove informative with this respect.

11.2   Sources for HL variability

In order to see which predictive variables explain this variability, another mixed effects regression model was calculated. We entered *participant* and *evidentiality type* as random variables in the baseline model as the previous analyses above already showed the effect of evidentiality type within and between groups.

While building the complete model each predictor was entered individually first. This includes the previously mentioned age-specific *L1/L2 richness, L1 experience* and *current L1 contact* variables. The predictors that reached significance alone or in interaction were then entered one by one to see whether they would improve

the simpler model. This decision was made based on the decrease in the Akaike Information Criterion (AIC) levels and p-values were obtained by likelihood ratio tests (Baayen et al. 2008). The model was based on 2285 observations. Table 7 shows the coefficients for the fixed effect factors in the final model.

Table 7. Factors accounting for individual variability in the evidentiality accuracy

	Estimate	SE	z	P
(intercept)	6.071478	3.04095	1.996573	0.0459 *
L2 richness (0–5)	−2.57136	1.070205	−2.40268	0.016276 *
LI experience (0–5)	2.698724	2.882012	−0.9364	0.349066
L2 richness (0–5):Ll experience (0–5)	4.625606	1.827925	2.530523	0.011389 *

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

According to Table 7, as the negative estimate shows, HSs with a rich L2 English environment between the ages 0 and 5 ($\beta = -2.57$, $z = -2.40$) showed a higher tendency to use the direct evidential in indirect evidential contexts. The effect of *L1 experience (0–5)* was not significant. However, as the positive estimate of the interaction between L2 richness and L1 experience in the model ($\beta = 4.62$, $z = 2.53$) indicates, the negative effect of L2 richness was compensated by the effect of past L1 experience (0–5). More precisely, for those whose L2 environment was rich, this factor seems to be a strong moderator in that an increase in their L1 experience (input and output) decreased the effect of the L2 richness.

None of the other predictive L1/L2 variables or their interaction with other variables contributed significantly to the model. Given this, in the following sections all references to past L1 experience and L1/L2 richness factors will pertain to this age range (0–5) unless specified otherwise.

11.3 Individual analysis

Given the large amount of variability between the HSs, we divided them into two groups: one that performed within the control range (native-like performers, henceforth NPs, $n = 10$) and those who fell outside that range (non-native-like performers: NNPs, $n = 21$).

As Table 8 shows, the individual examination of the NPs showed that they had been exposed to a slightly larger amount of L1 input ($M = 83\%$) during the ages 0–5, had had a slightly richer L1 environment ($M = 1.43$) during the same period, and were using their L1 interactively more frequently ($M = 82\%$) in comparison to the NNPs. Despite its lack of contribution to the model, interactive L1 use (with parents, relatives) was a significant predictor alone ($\beta = 5.55$, $SE = 2.05$, $z = 2.71$, $p = 0.006$).

In contrast, the L2 environment of the NNPs during the ages 0–5 had been slightly richer (*M* = 1.03), and they were using their L1 interactively slightly less frequently (75%) with their parents, relatives etc. but more frequently with their own HS friends (48%) in comparison to the NPs. The *L1 use with friends* variable however, was not a statistically explanatory variable.

All the rest of the variables' mean scores after the age of 5 were similar across the groups as can be seen in Table 8. One exception to this was L2 richness (5–11) and (11–18) interestingly in favor of the NPs. This would explain why these variables did not contribute to the regression model above.

Table 8. Mean scores of the individual analysis

Time		NP			NNP		
		Mean	Range	SD	Mean	Range	SD
	LI experience (%)	0.83	0.67–0.93	0.09	0.68	0.25–0.96	0.2
0–5	LI richness (out of 5)	1.43	0.84–2.1	0.39	1.1	0.00–2.84	0.68
	L2 richness (out of 5)	0.81	0.23–1.91	0.56	1.03	0.1–2.89	0.8
	LI experience (%)	0.41	0.27–0.57	0.12	0.38	0.16–0.65	0.14
5–11	LI richness (out of 5)	0.86	0.00–2.02	0.64	0.81	0.00–1.59	0.52
	L2 richness (out of 5)	1.87	1.48–2.29	0.28	1.59	0.6–3.47	0.71
	LI experience (%)	0.35	0.13–0.54	0.14	0.34	0.04–0.57	0.16
11–18	LI richness (out of 5)	1.07	0.00–2.82	0.81	1.01	0.00–2.38	0.66
	L2 richness (out of 5)	2.25	1.24–3.35	0.66	1.9	0.63–4.00	0.87
current	interactive LI use (%)	0.82	0.65–0.90	0.08	0.75	0.42–1.00	0.16
LI	LI passive exposure (%)	0.55	0.08–0.83	0.28	0.56	0.00–1.00	0.29
contact	LI use with friends (%)	0.33	0.13–0.53	0.13	0.48	0.13–1.00	0.26

12. Discussion

The present study aimed to investigate whether the potential divergent L1 grammar of adult Turkish HSs in the UK can be explained by qualitatively modified input conditions. With a widely-accepted assumption that the L1 would be transmitted to the next generation (HSs) via the first generation immigrants, the L1 performance of a group of first generation Turkish immigrants (LBs) in the UK and that of mono-linguals in Turkey (CG) was compared to see whether L1 input to be transmitted to the HSs is attrited. In order to see whether the L1 performance of the HSs would mirror that of the LBs or the CG, their performance was added into this comparison as well. A further interest was whether the HSs' overall accuracy in the evidentiality would relate to quantity and quality of their past and/or current L1 contact.

The first step was to look at the total number of finite verbs and the distribution of past/non-past usages across groups. The HSs used a slightly bigger number of finite verbs than both the LBs and the CG (per 1000 words). Although this difference was not significant ($F(2, 103) = 1.176$, $p = 0.313$), this tendency is not surprising given that previous findings have shown a tendency among HSs to avoid complex embedded clauses in Turkish (Treffers-Daller, Ozsoy & van Hout 2007) which might have caused an increase in the number of simple finite sentences or finite clauses (Onar Valk 2015). It is noted that Yılmaz (2013) did not find a similar tendency among first-generation Turkish migrants, except where the most complex embedding structures were concerned.

As for the frequency distribution of tenses, no significant differences were revealed for the marker of direct experience (dpast). The distribution of the inferential and reportative marker (mpast) however, was slightly in favour of the HSs (they used more mpast, which was a somewhat surprising finding). Although four LBs and two HSs did not use any repE forms at all and preferred the narrative present marker instead (an alternative form to narrate past events), this marker was not responsible for this inequality at the group level; an indication that none of the populations systematically avoided using indirect evidentials (mpast). The only remaining explanation seems to be the nature of the interview questions and different reactions of the participants to them. For the repE for example, all participants were asked to report childhood events as told by their parents. Given the HSs' younger age and that a great majority of them were still living with their parents at the time of testing, it was easy for them to talk about these fresh memories immediately. The same question elicited fewer memories from the LBs and older monolinguals, as in most cases they did not remember such events and had nothing to tell. Similarly, the "theft scenario" elicited longer reactions and more infE from the HSs as it was more appealing to the younger participants than the older ones. The LBs and many CG participants were more reluctant to role play with the interviewer and thus produced fewer infE. Future studies should take these effects into consideration when designing their elicitation tasks.

While the CG and the LBs performed at ceiling for all types of evidentials, this was not the case for the HSs. The error analysis determined that the HSs made a significant number of contextually inappropriate substitutions indicating a non-target-like L1 grammar, similar to what has often been found among HSs (e.g. Montrul 2008, 2016a).

Based on Rothman's model, the deviant forms found among HSs should mirror a non-target-like L1 performance of the LBs (due to attrition). This would have indicated that the HSs had complete acquisition of an attrited variety (Pascual y Cabo & Rothman 2012). In the present study this was not found to be the case: there was

no difference between the LBs and CG speakers at the group level, and only three LBs fell outside the control range. This indicates that the L1 input provided by this group of LBs can be considered qualitatively native-like in terms of the evidential structures. Our findings therefore do not support the qualitatively modified input claim (Rothman 2009), at least for this property. The LBs (all AoA > 12) seem to have maintained their L1 despite 18.7 years average length of residence (LOR) which is in line with the extensive amount of previous research that revealed no significant effects of external factors on L1 attrition once it starts upon full acquisition of the L1 (e.g. Schmid 2011).

This finding brings us back to a critical question: what are the sources of this divergent L1 grammar if the input is qualitatively native-like? This might be better approached with a more detailed look at the nature of the divergences. As evidenced by the analyses, the HSs substituted the repE more than they substituted the infE with the Dexp. However, with one single exception, no substitutions were made the other way. None of the participants replaced any evidentials with any non-past tense markers indicating awareness of the anteriority for both direct and indirect evidentials.

It is striking that even the least target-like speakers managed to use indirect evidentials context-appropriately (to some limited degree) and did not avoid the forms by overusing the narrative present marker. This shows their awareness of how the indirect evidentials' semantic requirements apply to certain contexts. Nevertheless, the large number of replacements with the Dexp suggests that the notion of tense was maintained to a greater degree than the notion of evidentiality (witnessing versus non-witnessing distinction) in the L1 of these HSs. This agrees with the observations that Spanish HSs maintained the notion of tense and aspect more than they did the modal categories (Montrul 2009), and that Turkish HSs were less sensitive to evidentiality violations than they were to time reference violations in non-evidential contexts (Arslan, De Kok & Bastiaanse 2015).

As also proposed by Arslan and Bastiaanse (2014, cited in Arslan, Bastiaanse & Felser 2015), the replacement tendencies show that the evidentiality status (visual access) of the direct evidential is affected in a way that enables the participants to extend its meaning to non-witnessed contexts. However, allowing the Dexp in non-witnessed contexts additionally indicates that the special context requirements of the indirect evidentials are also affected. Our observation thus seems to be perfectly compatible with what Arslan, de Kok and Bastiaanse (2015) revealed for decreased sensitivity to indirect evidential (repE) violations in Turkish-Dutch adult HSs. However, it is not entirely in line with what Arslan, Bastiaanse and Felser (2015) found in their study where the online processing of indirect evidentials (infE) was intact in the grammar of Turkish-German adult HSs. It should, however, be noted that the finding of the latter study shows participants' passive command

of the infE that does not necessarily lead to context-appropriate production. An investigation of the processing of the repE, which seems to be affected more than the infE in the current study and that of Arslan, de Kok and Bastiaanse (2015), might yield different results.

Unlike what has been suggested in this previous work, the fact that the HSs in the current study never used indirect evidentials in direct evidential contexts does not necessarily indicate that the semantic and pragmatic components of indirect evidentials were all retained. Rather, it points to "unstable knowledge" of witnessed versus non-witnessed distinction in the Turkish past tense. Whether this could be an outcome of incomplete acquisition or attrition (as it is not of qualitatively different input) is open to discussion. Evidentiality is a relatively late acquired property which does not stabilize before the ages of 5–6 (Aksu-Koç 1988; Ozturk & Papafragou 2008a among them). Given the average AoA of the participants (2.8 years) is just around the time indirect evidentials start appearing, input interruptions due to an early AoA might have caused a failure in developing age-appropriately. This might have resulted in incomplete L1 knowledge in adulthood due to insufficiency of the continuous L1 input (Montrul 2008). One third of the participants, however, managed to achieve and maintain native-like proficiency in the indirect evidentials despite an early AoA and interrupted L1 input.

Given the above factors, attrition might be at play as well. Some participants might have acquired the property despite the interrupted input, and experienced attrition in later stages possibly due to reduced L1 contact. Based on the revealed acquisition order of the evidentials, the HSs seem to have shown a reverse order of acquisition in their maintenance reminiscent of Jakobson's Regression Hypothesis (1941). The regression hypothesis would imply a complete acquisition of the evidentials which then might have started attriting. As discussed below, our data might not be able to fully answer whether this is the case unless the same property is examined longitudinally.

The lack of a longitudinal design in the current study would not fully allow us to disentangle incomplete acquisition from attrition. However, given the detailed past and current L1 contact data, both the statistical and the individual analysis results allow us to examine this distinction in more detail. These analyses revealed that the participants with a rich L2 environment were less accurate in evidentials. However, an increased amount of L1 exposure counteracted this L2 richness effect leading to more accurate usages. As the individual analysis unveiled, after the age of 5, L1 experience and L1 richness mean scores (5–18) did not differ much between the NPs and the NNPs. It is thus likely that the NNPs' less frequent early L1 experience, along with the slightly richer L2 environment, hindered them from developing age-appropriate forms. Their continuous L1 experience was presumably not sufficient for this development after this age either. It is therefore likely that the

NNPs did not attrite between the ages of 5–18 as their development was presumably never age-appropriate.

The NPs maintained this property even if their environment during school years (5–18) was much richer in terms of L2 compared to that of the NNPs. This indicates the importance of L1 experience in particular during the early years of linguistic development, both to acquire (Unsworth et al. 2014) and maintain the L1 long-term (Kondo-Brown 2005). It seems that the ultimate attainment, which looks "incomplete" at the surface level, is not a result of an early "AoA" or "interrupted L1 input". It rather appears to result from the fact that the amount of L1 experience remained insufficient to resist the effect of a rich L2 environment. It seems that there is a critical amount of input/output (Aksu-Koç et al. 2014) of around 83% (see Table 8) necessary to acquire the property during the primary years and resist the effect of the L2 richness, which was not reached by the NNPs either during the early years or later.

Given the NNPs' increased current L1 contact (48%) with other HSs in comparison to that of the NPs (33%), interaction with less proficient HSs might have triggered some sort of attrition in later stages resulting in even less-target-like L1 grammar than they had until the age of 18 (Montrul 2016c). However, this variable did not predict the outcome. It seems that neither incomplete acquisition nor attrition alone is enough to fully account for the findings.

At this point, a relatively recent model proposed by Putnam and Sánchez (2013) might be more explanatory. According to this model,[12] the HL acquisition is not interrupted or incomplete (see also Pascual y Cabo & Rothman 2012) but a continuum involving exposure shifts to L1 and L2 lexical items. These exposure shifts result in different levels of activation of functional features (FFs) depending on language use and processing patterns (rather than the amount of exposure) in both the L1 and L2. Lower activation levels for the production purposes (infrequent language use) would result in the features' becoming less available to retrieve and thus might be responsible for some L1 features to be replaced by FFs that are more activated in the L2, leading to a feature reassembly (Lardiere 2008). Considering

---

12. In this model the main focus is on the process of language acquisition as a whole rather than the outcome only, unlike incomplete acquisition and attrition accounts. Adopting a generative perspective for the role of input in acquisition, they reject the dependency of HL acquisition solely on the quantity and quality of input. On this view, a language can be acquired from a low frequency of input as well. The acquisition would rather depend on the "frequency of processing for comprehension and production purposes", i.e. intake and activation (Putnam & Sanchez 2013: 480). This partly contradicts with the main assumptions and findings of the current study about the role of input quantity for HL acquisition/maintenance. Nevertheless, some premises of the approach seem to be promising in order to explain the development of the L1 over the years as well as the degree of the maintenance observed.

our data from such a perspective, access to L1 lexical items with L1 use on a more frequent basis than the L2 in the early years would result in activation of L1 formal features. Continued L1 activation after this age (5–18), even if on a less frequent basis than the L2, would contribute to the maintenance of the L1 FFs long-term. In that sense, this approach is promising to account for the variability in the performance of our participants. For example, as acknowledged by the authors as well, it is very likely for repE FFs, being the last to be acquired, not to have ever been activated enough in the case of the NNPs due to more intense L2 use and activation. This would be in agreement with what we have concluded above about the role of early L1 experience (access and activation) in acquiring the property. As L1 use continued for the NNPs at almost an equal amount to the NPs, perhaps it was not intense enough (especially if L1 experience involved code-switching) (Putnam & Sánchez 2013) to keep the FFs activated. Similarly, more frequent interactive current L1 use by the NPs (see Table 8) might have contributed to FF activation in adulthood. For the NNPs on the other hand, increased L1 use with other HS friends would encourage code-switching and L1 use in bilingual mode (Schmid 2007). This may have affected the availability of the features. Features that become less available in time might have caused failure in form-meaning mappings leading to a dissociation and reassembly by the features in the L2.[13]

Both Turkish and English have the functional category [+/−past]. Turkish past tense however, additionally includes evidentiality morphology distinguishing between witnessed and unwitnessed past. From a feature reassembly point of view, the replacement errors observed would be a result of mapping L2 English [−evidentiality] values (because it is more dominant and accessed) onto L1 items causing them to undergo a dissociation and restructuring of [+evidentiality] values (Putnam & Sánchez 2013). This evaluation closely matches the conclusions derived before about the changes in the past tense system in the L1 of the HSs. Summarizing

---

13. As pointed out by one of the reviewers, the way Putnam and Sánchez's model explains how L1 grammar is changed ultimately due to the effect of a more activated L2 resembles how language change, in general, is approached by usage-based accounts. While the former describes this outcome as *feature reassembly*, it would be described as *contact-induced grammaticalization* by the latter (e.g. Heine & Kuteva 2005). Although these two accounts are clearly distinguished in terms of how they explain language acquisition in the first place (innate versus domain-general learning mechanisms) and the role of input quantity in language development, a detailed comparison between the premises of these accounts and how they account for L2-induced changes might inform our understanding more in the future. Note that the main reason why the current study consulted Putnam and Sánchez's model rather than other contact-induced explanations is primarily because the former has been explicitly developed as a reaction to the term „incomplete acquisition" and has specific predictions for non-convergent HL grammars which the current study was designed to address.

these findings, the extension of the meaning of the Dexp to nonwitnessed contexts highlights two points:

1. The evidential meaning of the Dexp was reassembled while its anteriority meaning was preserved. The reassembly was presumably due to its surface similarity to the English past tense marker and lower levels of activation.
2. This lower level of activation affected the availability of the indirect evidential values to be used in required contexts leading to a failure in form-meaning mappings. It is also likely that they were not activated enough during the linguistic development due to fluctuations in the activation levels of L1 and L2 FFs.

As claimed by Putnam and Sánchez (2013: 481) grammatical features that are considered to be a result of incomplete acquisition or attrition in the grammar of HSs are indeed a result of a failure in "mapping these features together in ways that are expected/predicted in monolingual variants of the heritage language".

In summary, the Putnam and Sánchez (2013) model looks at the phenomenon from a wider perspective and provides a more detailed explanation to HL divergences in comparison to incomplete acquisition and attrition accounts. In line with one of the premises of this model, the detailed past and current L1 experience data allowed us to confirm that HL divergences are not due to "interrupted input". One thing that this model does overlook but the current study captures, however, is that some L1 FFs might have remained underspecified not only because of more activated FFs in the L2 but also due to reduced amount of L1 input in the first place as emphasized by the incomplete acquisition and attrition accounts.

13. Concluding remarks

The distinction between different sources of information in Turkish past tense seems to be unstable in HL grammars due to a more activated L2 leading to a reassembled simplified L1 grammar as well as insufficient input received during the primary linguistic developmental stage. Not supporting the qualitatively different input explanation (Rothman 2007) as a possible cause of this, the results have been discussed from other approaches' point of views.

Interaction between two languages is indeed very complex. None of the approaches proposed to account for HL divergences are in fact capable of explaining this outcome alone: they could all contribute to the HL divergences to various degrees depending on several factors (Bayram, Pascual y Cabo & Rothman 2019). For example, not being able to provide evidence for the qualitatively different

input claim does not mean that it is not a contributing factor to HL divergences for other properties or for the performance of subsequent generations. In fact, the role of heritage speakers in diachronic change and appearance of new varieties has been the centre of focus in contact linguistics as these speakers seem to lead the change (Onar Valk & Backus 2013). To what extent this reassembly process, which includes features from both L1 and L2 described above, or contact-induced grammaticalization (as the contact linguistics would describe it, e.g. Heine & Kuteva 2005) leads to a new variety (permanent changes) is yet to be determined. Future research should focus on the L1 performance of subsequent generations in terms of evidentiality to see whether this new simplified L1 past tense would be reflected in the performance of children of these HSs and perhaps becomes the norm in immigrant Turkish over the years.

Abbreviations

1	first person	GEN	genitive case
2	second person	LOC	locative case
3	third person	M.PAST	*mIş* past tense
ABL	ablative case	PASS	passive
ACC	accusative case	PL	plural
AOR	aorist	POSS	possessive
CV	converb	SG	singular.
D.PAST	*DI* past tense		

References

Aarssen, J. 2001. Development of temporal relations in narratives by Turkish-Dutch bilingual children. In *Narrative Development in a Multilingual Context* [Studies in Bilingualism 23], L. Verhoeven & S. Strömqvist (eds), 209–232. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.23.08aar

Ågren, M., Granfeldt, J. & Thomas, A. 2014. Combined effects of age of onset and input on the development of different grammatical structures: A study of simultaneous and successive bilingual acquisition of French. *Linguistic Approaches to Bilingualism* 4(4): 462–493. https://doi.org/10.1075/lab.4.4.03agr

Aikhenvald, A. Y. 2004. *Evidentiality*. Oxford: OUP.

Aksu-Koç, A. 1988. *The Acquisition of Aspect and Modality: The Case of Past Reference in Turkish*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511554353

Aksu-Koç, A. 1994. Development of linguistic forms: Turkish. In *Relating Events in Narrative: A Crosslinguistic Developmental Study*, R. A. Berman & D. I. Slobin (eds), 329–385. Mahwah NJ: Lawrence Erlbaum Associates.

Aksu-Koç, A. 2000. Some aspects of the acquisition of evidentials in Turkish. In *Evidentials: Turkic, Iranian and Neighbouring Languages*, L. Johanson & B. Utas (eds), 15–28. Berlin: De Gruyter. https://doi.org/10.1515/9783110805284.15

Aksu-Koç, A. 2009. Evidentials: An interface between linguistic and conceptual development. In *Crosslinguistic Approaches to the Psychology of Language: Research in the Tradition of Dan Isaac Slobin*, J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura & S. Ozcalisk (eds), 531–542. London: Taylor & Francis.

Aksu-Koç, A., Ogel-Balaban, H. & Alp, I. E. 2009. Evidentials and source knowledge in Turkish. *New Directions for Child and Adolescent Development* 2009(125), 13–28. https://doi.org/10.1002/cd.247

Aksu-Koç, A., Terziyan, T. & Erguvanlı-Taylan, E. 2014. Input offers and child uptakes: Acquiring mood and modal morphology in Turkish. *Language Interaction and Acquisition* 5(1): 62–81. https://doi.org/10.1075/lia.5.1.03aks

Albirini, A., Benmamoun, E. & Saadah, E. 2011. Grammatical features of Egyptian and Palestinian Arabic heritage speakers' oral production. *Studies in Second Language Acquisition* 33: 273–303. https://doi.org/10.1017/S0272263110000768

Arslan, S. & Bastiaanse, R. (2014). Tense/aspect and evidentiality in narrative speech of Turkish-Dutch bilingual speakers. Paper presented at the 17th International conference on Turkish linguistics, University of Rouen, France.

Arslan, S., Bastiaanse, R. & Felser, C. 2015. Looking at the evidence in visual world: eye-movements reveal how bilingual and monolingual Turkish speakers process grammatical evidentiality. *Language Sciences* 1387. https://doi.org/10.3389/fpsyg.2015.01387

Arslan, S., De Kok, D. & Bastiaanse, R. 2017. Processing grammatical evidentiality and time reference in Turkish heritage and monolingual speakers. *Bilingualism: Language and Cognition* 20(3): 457–472. https://doi.org/10.1017/S136672891500084X

Baayen, R. H., Davidson, D. J. & Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4): 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bacanlı, E. 2006. Türkçedeki Dolaylılık İşaretleyicilerin Pragmatik Anlamları. *Modern Türklül Araştırmaları Dergisi* 1(3): 35–47.

Bacanlı, E. 2008. Türkiye Türkçesindeki -miş Ekinin Dolaylılık ve Dolaylılık-Dışı Kullanımlarında Zamansal Atıf. *Bilig* 44: 1–24.

Bates, D., Mächler, M., Bolker, B. & Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1–48. https://doi.org/10.18637/jss.v067.i01

Bayram, F., Pascual y Cabo, D. & Rothman, J. 2019. Intra-generational attrition contributions to heritage speaker competence. In *The Oxford Handbook of Language Attrition*, M. S. Schmid & B. Köpke (eds), 446–457. New York: Oxford University Press.

Benmamoun, E., Montrul, S. & Polinsky, M. 2013. Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics* 39(3–4): 129–181. https://doi.org/10.1515/tl-2013-0009

Bylund, E. 2009. Maturational constraints and first language attrition. *Language Learning* 59(3): 687–715. https://doi.org/10.1111/j.1467-9922.2009.00521.x

Çoban, C. 2014. İstanbul'da bir binanın üzerine vinç devrildi. *Hürriyet 30 November 2014*. <http://www.hurriyet.com.tr/istanbulda-bir-binanin-uzerine-vinc-devrildi-27670596> (3 January 2015).

Coşkun, H. 2010. Embedding indirective (evidential) utterances in Turkish. In *Linguistic Realization of Evidentiality in European Languages*, G. Diewald, E. Smirnova (eds), 187–221. Berlin: De Gruyter.

Csato, E. A. 2000. Turkish MIş- and IMIş- items: Dimensions of a functional analysis. In *Evidentials: Turkic, Iranian and Neighbouring Languages*, L. Johanson & B. Utas (eds), 29–44. Berlin: De Gruyter. https://doi.org/10.1515/9783110805284.29

De Houwer, A. 2007. Parental language input patterns and children's bilingual use. *Applied Psycholinguistics* 28(3): 411–424. https://doi.org/10.1017/S0142716407070221

De Houwer, A. 2009. *An Introduction to Bilingual Development*. Bristol: Multilingual Matters.

DeLancey, S. 2001. The mirative and evidentiality. *Journal of Pragmatics* 33(3): 369–382. https://doi.org/10.1016/S0378-2166(01)80001-1

Gaziantep'te sel felaketi. 2014. *Milliyet 15 October 2014*. <http://www.milliyet.com.tr/gaziantep-te-sel-felaketi-gaziantep-yerelhaber-425385> (3 January 2015).

Göksel, A. & Kerslake, C. 2005. *Turkish: A Comprehensive Grammar*. London: Routledge. https://doi.org/10.4324/9780203340769

Gollan, T. H., Starr, J. & Ferreira, V. S. 2014. More than use it or lose it: The number-of-speakers effect on heritage language proficiency. *Psychonomic Bulletin & Review* 22(1): 147–155. https://doi.org/10.3758/s13423-014-0649-7

Gries, S. T. 2015. The most under-used statistical method in corpus linguistics: multi-level (and mixed-effects) models. *Corpora* 10(1): 95–125. https://doi.org/10.3366/cor.2015.0068

Gutierrez-Clellen, V. F. & Kreiter, J. 2003. Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics* 24(2): 267–288. https://doi.org/10.1017/S0142716403000158

Heine, B. & Kuteva, T. 2005. *Language Contact and Grammatical Change*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511614132

Hoff, E. 2006. How social contexts support and shape language development. *Developmental Review* 26(1): 55–88. https://doi.org/10.1016/j.dr.2005.11.002

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M. & Parra, M. 2012. Dual language exposure and early bilingual development. *Journal of Child Language* 39(1), 1–27. https://doi.org/10.1017/S0305000910000759

Iglowstein, I., Jenni, O. G., Molinari, L. & Largo, R. H. 2003. Sleep duration from infancy to adolescence: Reference values and generational trends. *Pediatrics* 111(2): 302–307. https://doi.org/10.1542/peds.111.2.302

Jakobson, R. 1941. *Kindersprache, Aphasie, and allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell.

Jia, R. & Paradis, J. 2014. The use of referring expressions in narratives by Mandarin heritage language children and the role of language environment factors in predicting individual differences. *Bilingualism: Language and Cognition, First View*, 1–16. https://doi.org/10.1017/S1366728914000728

Johanson, L. 2000. Turkic indirectives. In *Evidentials: Turkic, Iranian and Neighbouring Languages*, L. Johanson & B. Utas (eds), 61–88. Berlin: De Gruyter. https://doi.org/10.1515/9783110805284.61

Johanson, L. 2006. Indirective sentence types. *Turkic Languages* 10: 73–89.

Kaltsa, M., Tsimpli, I. M. & Rothman, J. 2015. Exploring the source of differences and similarities in L1 attrition and heritage speaker competence: Evidence from pronominal resolution. *Lingua* 164B: 266–288. https://doi.org/10.1016/j.lingua.2015.06.002

Karakoç, B. 2007. Connectivity by means of finite elements in monolingual and bilingual Turkish discourse. In *Connectivity in Grammar and Discourse* [Hamburg Studies on Multilingualism 5], J. Rehbein, C. Hohenstein & L. Pietsch (eds), 199–228. Amsterdam: John Benjamins. https://doi.org/10.1075/hsm.5.13kar

Kondo-Brown, K. 2005. Differences in language skills: heritage language learner subgroups and foreign language learners. *The Modern Language Journal* 89(4): 563–581. https://doi.org/10.1111/j.1540-4781.2005.00330.x

Kupisch, T. & Rothman, J. 2018. Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism* 22(5): 564–582. https://doi.org/10.1177/1367006916654355

La Morgia, F. 2011. Bilingual First Language Acquisition: The Nature of the Weak Language and the Role of the Input. PhD dissertation, Dublin City University. <http://doras.dcu.ie/16056/> (17 October 2014).

Lardiere, D. 2008. Feature-assembly in second language acquisition. In *The Role of Formal Features in Second Language Acquisition*, J. Liceras, H. Zobl & H. Goodluck (eds), 106–140. New York NY: Lawrence Erlbaum Associates.

Lazard, G. 2001. On the grammaticalization of evidentiality. *Journal of Pragmatics* 33(3): 359–367. https://doi.org/10.1016/S0378-2166(00)00008-4

Leseman, P. P. M., Scheel, A. F., Mayo, A. Y. & Messer, M. H. 2009. Bilingual development in early childhood and the languages used at home: Competition for scarce resources? *Streitfall Zweisprachigkeit – the Bilingualism Controversy*, I. Gogolin & U. Neumann (eds), 289–316. Wiesbaden: Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91596-8_17

Long, D. & Rothman, J. 2014. Some caveats to the role of input in the timing of child bilingualism. *Linguistic Approaches to Bilingualism* 4(3): 351–356. https://doi.org/10.1075/lab.4.3.09lon

Lytra, V. 2012. Discursive constructions of language and identity: Parents' competing perspectives in London Turkish complementary schools. *Journal of Multilingual and Multicultural Development* 33(1): 85–100. https://doi.org/10.1080/01434632.2011.638076

MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk*, 3rd edn. Mahwah NJ: Lawrence Erlbaum Associates.

Meriç suyunun hapsettiği mahalle: Karaağaç. 2014. *Milliyet 9 December 2014.* <http://www.milliyet.com.tr/meric-suyunun-hapsettigi-mahalle--gundem-1981678> (3 January 2015).

Mindell, J. A., Sadeh, A., Wiegand, B., How, T. H. & Goh, D. Y. T. 2010. Cross-cultural differences in infant and toddler sleep. *Sleep Medicine* 11(3): 274–280. https://doi.org/10.1016/j.sleep.2009.04.012

Montrul, S. 2008. *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor* [Studies in Bilingualism 39]. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.39

Montrul, S. 2009. Knowledge of tense-aspect and mood in Spanish heritage speakers. *International Journal of Bilingualism* 13(2): 239–269. https://doi.org/10.1177/1367006909339816

Montrul, S. 2010. Current issues in heritage language acquisition. *Annual Review of Applied Linguistics* 30: 3–23. https://doi.org/10.1017/S0267190510000103

Montrul, S. 2016a. *The Acquisition of Heritage Languages*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139030502

Montrul, S. 2016b. Heritage language development: Connecting the dots. *International Journal of Bilingualism*. Online first. https://doi.org/10.1177/1367006916654368

Montrul, S. 2016c. Language attrition and language transmission. Paper presented at the *Third International Conference on Language Attrition (ICLA3)*, Colchester, UK. <https://language-attrition.org/conference-program> (5 July 2016).

Montrul, S., Bhatt, R. M. & Bhatia, A. 2012. Erosion of case and agreement in Hindi heritage speakers. *Linguistic Approaches to Bilingualism* 2(2): 141–176. https://doi.org/10.1075/lab.2.2.02mon

Montrul, S., Davidson, J., Fuente, I. D. L. & Foote, R. 2014. Early language experience facilitates the processing of gender agreement in Spanish heritage speakers. *Bilingualism: Language and Cognition* 17(1): 118–138. https://doi.org/10.1017/S1366728913000114

Montrul, S. & Sánchez-Walker, N. 2013. Differential object marking in child and adult Spanish heritage speakers. *Language Acquisition* 20(2): 109–132. https://doi.org/10.1080/10489223.2013.766741

Mueller-Gathercole, V. C. 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. https://doi.org/10.2167/beb442.0

O'Grady, W., Lee, O.-S. & Lee, J.-H. 2011. Practical and theoretical issues in the study of heritage language acquisition. *Heritage Language Journal* 8(3): 23–40.

Olds, T., Blunden, S., Petkov, J. & Forchino, F. 2010. The relationships between sex, age, geography and time in bed in adolescents: A meta-analysis of data from 23 countries. *Sleep Medicine Reviews* 14(6): 371–378. https://doi.org/10.1016/j.smrv.2009.12.002

Onar Valk, P. 2015. Transformation in Dutch Turkish subordination? *Converging Evidence of Change Regarding Finiteness and Word Order in Complex Clauses*. PhD dissertation, Tilburg University, Utrecht. <https://pure.uvt.nl/portal/en/publications/transformation-in-dutch-turkish-subordination(6fde9f98-96bf-4765-ad8d-53909dd72f68).html> (9 November 2017).

Onar Valk, P. & Backus, A. 2013. Syntactic change in an immigrant language: from non-finite to finite subordinate clauses in Turkish. *Journal of Estonian and Finno-Ugric Linguistics* 4(2): 7–29. https://doi.org/10.12697/jeful.2013.4.2.01

Ozturk, O. & Papafragou, A. 2008a. The acquisition of evidentiality and source monitoring. In *Proceedings from the 32nd Annual Boston University Conference on Language Development*, H. Chan, H. Jacob & E. Kapia (eds), 199–230. Somerville MA: Cascadilla Press.

Ozturk, O. & Papafragou, A. 2008b. The acquisition of evidentiality in Turkish. *University of Pennsylvania Working Papers in Linguistics* 14(1): 297–309.

Ozturk, O. & Papafragou, A. 2016. The acquisition of evidentiality and source monitoring. *Language Learning and Development* 12(2): 199–230. https://doi.org/10.1080/15475441.2015.1024834

Paradis, J. 2011. Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism* 1(3): 213–237. https://doi.org/10.1075/lab.1.3.01par

Paradis, J. & Genesee, F. 1996. Syntactic Acquisition in Bilingual Children. *Studies in Second Language Acquisition* 18(1): 1–25. https://doi.org/10.1017/S0272263100014662

Pascual y Cabo, D. 2013. Agreement Reflexes of Emerging Optionality in Heritage Speaker Spanish. PhD dissertation, University of Michigan, Ann Arbor. <http://search.proquest.com/docview/1521060739/abstract/BE4FD3FCCBB64266PQ/1> (10 May 2017).

Pascual y Cabo, D. & Rothman, J. 2012. The (il)logical problem of heritage Speaker bilingualism and incomplete acquisition. *Applied Linguistics* 33(4): 450–455. https://doi.org/10.1093/applin/ams037

Pires, A. & Rothman, J. 2009. Disentangling sources of incomplete acquisition: An explanation for competence divergence across heritage grammars. *International Journal of Bilingualism* 13(2): 211–238. https://doi.org/10.1177/1367006909339806

Polinsky, M. 2011. Reanalysis in adult heritage language. *Studies in Second Language Acquisition* 33(2): 305–328. https://doi.org/10.1017/S027226311000077X

Putnam, M. T. & Sánchez, L. 2013. What's so incomplete about incomplete acquisition? A prolegomenon to modeling heritage language grammars. *Linguistic Approaches to Bilingualism* 3(4): 478–508. https://doi.org/10.1075/lab.3.4.04put

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/> (19 July 2016).

Rothman, J. 2007. Heritage speaker competence differences, language change, and input type: Inflected infinitives in Heritage Brazilian Portuguese. *International Journal of Bilingualism* 11(4): 359–389. https://doi.org/10.1177/13670069070110040201

Rothman, J. 2009. Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism* 13(2): 155–163. https://doi.org/10.1177/1367006909339814

Scheele, A. F., Leseman, P. P. M. & Mayo, A. Y. 2010. The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics* 31(1): 117–140. https://doi.org/10.1017/S0142716409990191

Schmid, M. S. 2007. The role of L1 use for L1 attrition. In *Language Attrition: Theoretical Perspectives* [Studies in Bilingualism 33], B. Köpke, M. S. Schmid, M. Keijzer & S. Dostert (eds), 135–153. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.33.10sch

Schmid, M. S. 2011. *Language Attrition*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511852046

Schmid, M. S. & Dusseldorp, E. 2010. Quantitative analyses in a multivariate study of language attrition: the impact of extralinguistic factors. *Second Language Research* 26(1): 125–160. https://doi.org/10.1177/0267658309337641

Schroeder, C. 2000. Between resultative, historical and inferential:non-finite -mIş forms in Turkish. In *Evidentials: Turkic, Iranian and Neighbouring Languages*, B. Utas & L. Johanson (eds), 115–143. Berlin: De Gruyter. https://doi.org/10.1515/9783110805284.115

Şener, N. 2011. Semantics and Pragmatics of Evidentials in Turkish. PhD dissertation, University of Michigan, Ann Arbor. <http://search.proquest.com/docview/909527119/abstract?accountid=13014> (4 August 2014).

Sezer, E. 2001. Finite inflection in Turkish. In *The Verb in Turkish* [Linguistik Aktuell/Linguistics Today 44], E. Erguvanh Taylan (ed), 1–45. Amsterdam: John Benjamins.

Silva-Corvalán, C. 2016. Simultaneous bilingualism: Early developments, incomplete later outcomes? *International Journal of Bilingualism*. Online First. https://doi.org/10.1177/1367006916652061

Slobin, D. I. 2001. Form-function relations: How do children find out what they are? In *Language Acquisition and Conceptual Development*, M. Bowerman & S. C. Levinson (eds), 406–449. Cambridge: CUP. https://doi.org/10.1017/CBO9780511620669.016

Slobin, D. I. & Aksu, A. 1982. Tense, aspect, and modality in the use of the Turkish evidential. In *Tense-Aspect: Between Semantics and Pragmatics* [Typological Studies in Language 1], P. J. Hopper (ed), 185–200. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.1.13slo

Sorace, A. 2014. Input, timing, and outcomes in a wider model of bilingualism. *Linguistic Approaches to Bilingualism* 4(3): 377–380. https://doi.org/10.1075/lab.4.3.14sor

Tafolar, M. 2014. CHP, Yalova'daki ağaç kesimini incelemeye aldı. *Milliyet 1 December 2014*. <http://www.milliyet.com.tr/chp-yalova-daki-agac-kesimini-gundem-1977458> (1 December 2014).

Treffers-Daller, J., Daller, M., Furman, R. & Rothman, J. 2016. Ultimate attainment in the use of collocations among heritage speakers of Turkish in Germany and Turkish–German returnees. *Bilingualism: Language and Cognition* 19(3): 504–519.
https://doi.org/10.1017/S1366728915000139

Treffers-Daller, J., Ozsoy, A. S. & van Hout, R. 2007. (In)complete acquisition of Turkish among Turkish-German bilinguals in Germany and Turkey: An analysis of complex embeddings in narratives. *International Journal of Bilingual Education and Bilingualism* 10(3): 248–276.
https://doi.org/10.2167/beb443.0

Tsimpli, I. M. 2014. Early, late or very late: Timing acquisition and bilingualism: A reply to peer commentaries. *Linguistic Approaches to Bilingualism* 4(3): 393–402.
https://doi.org/10.1075/lab.4.3.17tsi

Ünal, E. & Papafragou, A. 2013. Linguistic and conceptual representations of inference as a knowledge source. In *37th Annual Boston University Conference on Language Development*, S. Baiz, N. Goldman & R. Hawkes, 433–443. Boston MA: Cascadilla Press.

Ünal, E. & Papafragou, A. 2016. Production-comprehension asymmetries and the acquisition of evidential morphology. *Journal of Memory and Language* 89: 179–199.
https://doi.org/10.1016/j.jml.2015.12.001

Unsworth, S. 2013. Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition* 16(1): 86–110. https://doi.org/10.1017/S1366728912000284

Unsworth, S. 2016. *Bilingual Language Exposure Calculator (BILEC): Manual.* Nijmegen: Radboud University. <https://www.iris-database.org> (6 October 2016).

Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A. & Tsimpli, I. M. 2014. The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics* 35(4): 765–805. https://doi.org/10.1017/S0142716412000574

van Osch, B. & Sleeman, P. 2016. Spanish heritage speakers in the Netherlands: Linguistic patterns in the judgment and production of mood. *International Journal of Bilingualism*. Online First. https://doi.org/10.1177/1367006916654365

Verhoeven, L. 2004. Bilingualism and narrative construction. In *Relating Events in Narrative: Typological and Contextual Perspectives*, S. Strömqvist & L. Verhoeven (eds), 435–454. Mahwah NJ: Lawrence Erlbaum Associates. <http://repository.ubn.ru.nl/handle/2066/63895> (20 July 2015).

Yılmaz, G. 2013. *Bilingual Language Development among the First Generation Turkish Immigrants in the Netherlands.* Groningen: University of Groningen. <http://www.let.rug.nl/languageattrition/Papers/Yilmaz%20dissertation%202013.pdf> (26 April 2015).

Yürürken üzerine duvar yıkıldı. 2014. *Hürriyet 28 November 2014.* <http://www.hurriyet.com.tr/yururken-uzerine-duvar-yikildi-27671396> (28 November 2014).

Not in the mood

Frequency effects in heritage speakers' subjunctive knowledge

David Giancaspro
University of Richmond

Previous studies (e.g., van Osch & Sleeman 2016; Perez-Cortes 2016) have found that heritage speakers (HSs) of Spanish produce less lexically-selected subjunctive mood morphology than Spanish-dominant speakers. It remains unclear, however, whether the HSs' tendency to produce less subjunctive mood than Spanish-dominant speakers is attributable to representational differences (e.g., Montrul 2002, 2008), input quality differences (e.g., Pires & Rothman 2009; Pascual y Cabo & Rothman 2012), or as yet unidentified factors. The present study addresses this question by testing the effect of lexical frequency on advanced proficiency HSs' productive (Experiment 1) and receptive (Experiment 2) knowledge of lexically-selected subjunctive mood in Spanish. Results of Experiments 1 and 2 indicate that advanced proficiency HSs are both (a) highly accurate with subjunctive mood and yet also (b) significantly *less* accurate with lower frequency verbs. Given these findings, as well as the categorical subjunctive production of the Spanish-dominant bilingual control group, it is argued that HSs' differences from dominant speakers may be (partially) attributable to gaps in lexical, rather than morphosyntactic knowledge.

Keywords: heritage speakers, subjunctive mood, variability, frequency

1. Introduction

Heritage speakers (HSs) are "individuals from minority language groups who grow up exposed to a minority language in the home and the majority societal language" (Montrul 2016: 16). Perhaps the most commonly studied HSs are HSs of Spanish in the United States. Typically, though not necessarily (e.g., Leal-Méndez, Rothman & Slabakova 2015), HSs of Spanish exhibit heritage language (HL) knowledge which differs, quantitatively and/or qualitatively, from the knowledge demonstrated by

Spanish-dominant speakers,[1] e.g., first-generation immigrants. HSs differ *quantitatively* from Spanish-dominant speakers when they produce significantly less of a given HL "target" form. Montrul (2009), for example, found that advanced proficiency HSs of Spanish produced less subjunctive mood morphology (92.3%) in expected subjunctive contexts than Spanish-dominant speakers (100%) did. HSs differ *qualitatively* from dominant speakers of Spanish when they produce linguistic forms which are qualitatively distinct from the forms produced by Spanish-dominant speakers. Pascual y Cabo & Gómez Soler (2015), for example, found that some HSs produce "non-target" preposition-stranding in Spanish (e.g., *quién bailas con*, "who do you dance with?"), perhaps due to English influence, even though these forms are not produced by Spanish-dominant speakers in their study.

In both cases, it can be said that HSs' differences from Spanish-dominant speakers exemplify linguistic "losses in transmission", in the sense that HSs develop Spanish grammars which diverge from the grammatical systems "transmitted" to them by Spanish-dominant speakers via the primary input. In Montrul (2009), HSs innovate beyond the input by adapting or developing a Spanish grammar in which both indicative and subjunctive mood morphology are possible in traditionally obligatory subjunctive contexts. Similarly, in Pascual y Cabo & Gómez Soler (2015), HSs innovate by forming a more expansive Spanish grammar that allows for *both* pied-piping *and* preposition-stranding.

Though HSs of Spanish, as well as HSs of many other HLs, differ from HL-dominant speakers[2] (e.g., their primary input providers) in the vast majority of

---

1.   When I use the term "Spanish-dominant speakers", I am referring to native speakers of Spanish who were born and raised in a Spanish-speaking country before immigrating to the United States (and learning English) *after puberty*. Though these speakers are truly bilinguals and, in many cases, use English frequently in their day-to-day lives, their relatively later ages of acquisition of English typically ensure that they remain more comfortable or "dominant" in Spanish, rather than English, unlike heritage speakers of Spanish, who typically report English dominance. For more on the self-reported language dominance of first-generation immigrants and heritage speakers, see Sections 3 and 4.

2.   I use the term "HL-dominant speakers" instead of "Spanish-dominant speakers" in order to expand the scope of the discussion from differences between HSs of Spanish and their input providers to differences between HSs of *any language* and their input providers. The term "HL-dominant speakers", therefore, is simply an umbrella term referring to the first-generation immigrants who provide HSs *of any HL* with their primary input in that language. In a hypothetical study of heritage speakers of Russian in the US, for example, the first-generation immigrants from Russia, against whom the HSs are compared, would be "HL-dominant speakers", in the sense that they are dominant in a language (Russian) that is the *heritage language* of the HSs. The HL-dominant speakers, however, have no HL. Throughout the remainder of the paper, I will use "HL-dominant speakers", rather than "Spanish-dominant speakers", when commenting on linguistic patterns/explanations relevant for not just Spanish but also any heritage language.

HL acquisition studies (see Kupisch & Rothman 2018 for discussion of a few notable exceptions), the underlying causes of these differences are still a source of great controversy among HL-acquisition researchers. How can we explain the fact that HSs so often differ quantitatively and/or qualitatively from their primary HL-input providers (e.g., first-generation immigrants)? In other words, how do some HL properties become "lost" (or changed) in transmission as the HL is passed from HL-dominant first generation speakers to HSs of the second and subsequent generations?

One prevalent account of HS differences is what I call the Representational Differences Approach (e.g., Montrul 2002, 2008), according to which HS differences with a given HL property are attributable to differences in HSs' *underlying representations* of that property. From this perspective, HSs who diverge quantitatively or qualitatively from HL-dominant speakers do so as a direct result of having acquired different underlying grammatical knowledge. Therefore, the HSs in Montrul (2009) produce less subjunctive mood in expected subjunctive contexts than HL-dominant speakers because they have different representations of mood in Spanish.

Though it is often found in the HL-acquisition literature, the Representational Differences Approach does not clearly explain the performance of HSs who, despite differing *quantitatively* from HL-dominant speakers, nonetheless maintain *qualitatively* similar grammatical knowledge. Take, for example, the advanced proficiency HSs in Montrul (2009), who produce subjunctive mood morphology in over 90% of expected subjunctive contexts. If these HSs have an underlyingly different (or "incomplete") representation of mood, as argued by Montrul, how do they manage to produce subjunctive nearly categorically in expected subjunctive contexts? Clearly, attributing *all* HS differences to underlyingly different HL representations makes it hard to understand how and why HSs produce target forms variably, yet well above chance.

A second prevalent account of HS differences is the Input Quality Approach (e.g., Pires & Rothman 2009; Pascual y Cabo & Rothman 2012), which highlights the possibility that some HS differences are directly attributable to the *input quality* that HSs receive from HL-dominant speakers. HL-dominant speakers, as they acquire the societal language and encounter contact-varieties of the HL, may begin to experience changes (or attrition) in their own HL knowledge, which they then pass on to HSs. In other words, some HS differences (e.g., producing less differential object marking (DOM) with animate, specific direct objects in Spanish, cf. Montrul & Sánchez-Walker 2013) are driven by changes taking place in the linguistic systems of the HL-dominant speakers, who expose HSs to those differences when providing them input in the HL.

It is undoubtedly important, following the Input Quality Approach, to consider the possibility that HS differences (e.g., producing less of a certain target form)

are the result of hearing qualitative differences in the input provided to them by HL-dominant speakers. Nonetheless, the explanatory scope of the Input Quality Approach is necessarily limited, as not all HS differences can be clearly attributed to input quality. HSs' production of preposition stranding in Spanish, for example, appears to emerge in HSs independent of input quality, given that HL-dominant speakers neither produce nor accept these forms, even after living for more than ten years in the US (Pascual y Cabo & Gómez Soler 2015). Consequently, there must be other factors beyond input quality which lead HSs to perform differently from HL-dominant speakers.

As illustrated thus far, neither the Representational Differences Approach nor the Input Quality Approach is sufficient to account for HS differences from HL-dominant speakers. It is critical, therefore, that researchers consider other, alternative factors underlying HS differences. The present study takes a step in this direction by investigating the effect of *lexical frequency* on HSs' knowledge of lexically-selected subjunctive mood morphology in Spanish. Though many studies have revealed a strong predictive role of lexical frequency in L2 acquisition (for a review, see Ellis 2002), there are very few studies which have coded for, let alone purposefully manipulated, the role of this critical variable in HL acquisition.

To my knowledge, only three studies in the HL acquisition literature directly suggest that lexical frequency may impact the extent to which HSs differ from HL-dominant speakers with grammatical properties of the HL. Dorian (1981) provides evidence that Gaelic HSs' variability with conditional verbal morphology may be at least somewhat tied to lexical frequency. One HS[3] in her study, most notably, was 100% accurate (3/3) in the production of conditional morphology on the highly frequent verb 'to be' but just 20% accurate (1/5) on other, less frequent Gaelic verbs. This finding raises the possibility that HSs' morphological divergence from HL-dominant speakers may be strongly conditioned by lexical frequency. Gal (1989) reports a similar finding from HSs of Hungarian living in Austria, who, unlike dominant speakers of Hungarian, appear to only use causative morphology with highly frequent Hungarian verbs. Neither of these studies carefully manipulated the role of lexical frequency on HSs' production of verbal inflections. Nonetheless, both studies suggest a potential effect of lexical frequency on HSs' HL divergence.

The most recent evidence of lexical effects on HSs' grammatical knowledge comes from Montrul, Davidson, De la Fuente, and Foote (2014), who tested HSs' knowledge of gender agreement with canonical and non-canonical nouns,[4]

---

3. In her book, Dorian uses the term "semi-speaker" to refer to what we would now call heritage speakers.

4. Canonical Spanish nouns are those nouns where the endings -o and -a faithfully indicate a noun's grammatical gender.

respectively. In all three experimental tasks, HSs performed well above chance in the recognition and production of gender agreement in Spanish, thereby exhibiting systematic knowledge of grammatical gender. Most interestingly, though, HSs performed significantly less accurately with non-canonical nouns in all three experimental tasks, thereby demonstrating that their productive and receptive knowledge of gender agreement is substantially affected by noun canonicity. To account for this finding, the authors suggest that HSs have acquired gender agreement but apply it less consistently to non-canonical nouns, due to reduced production of and exposure to these less frequent forms.

Given these suggestive findings, as well as HSs' well-documented (and currently unexplained) tendency to alternate between indicative and subjunctive mood forms in expected subjunctive contexts (e.g., Montrul 2009; Perez-Cortes 2016; van Osch & Sleeman 2018), the present study, which evaluates the effects of lexical frequency on HSs' subjunctive production, is a timely contribution to our understanding of HS differences in the domain of verbal morphology.

The remainder of this chapter is organized as follows. In Section 2, I will provide a brief overview of lexically-selected subjunctive mood and its acquisition by Spanish native speakers. In Section 3, I will present Experiment #1, which tests HSs' *productive* knowledge of lexically-selected subjunctive mood. In Section 4, I will present Experiment #2, which tests HSs' *receptive* knowledge of lexically-selected subjunctive mood. Finally, in Section 5, I will discuss the implications of the present findings as well as their relevance for future research on HL acquisition.

2. Lexically-subjunctive mood: Syntax and acquisition

2.1 Intensional subjunctive mood: Morphology and syntax

Many studies have shown that HSs differ from HL-dominant speakers in their productive and/or receptive command of HL verbal morphology (e.g., Montrul 2002, 2009; Polinsky 2006; Sherkina Lieber 2015). Of the many categories of verbal morphology (e.g., aspect, tense) which have been explored in HL acquisition research, mood morphology seems to be the "most affected" in HSs (Montrul 2016), making it an ideal starting point for a study of HS differences.

What is subjunctive mood morphology and what does it look like in Spanish? Broadly, subjunctive mood refers to a set of inflections used to mark modality (Bosque 2012). Subjunctive mood is expressed morphologically either by (a) a thematic vowel shift (e.g., *-e* to *-a*) or (b) a thematic vowel shift *and* a change in the verbal root (e.g., *dic-* to *dig-*). Table 1 summarizes these possibilities with verbs from both the *-ar* and *-er/-ir* verb classes, respectively.

Table 1. Instantiations of indicative and subjunctive mood in Spanish

	Infinitival form	Indicative form	Subjunctive form
I. Shift in Thematic Vowel	a. *comer* ('eat')	a. *com-e*	a. *com-a*
	b. *hablar* ('speak')	b. *habl-a*	b. *habl-e*
II. Shift in Thematic Vowel Plus Root Change	a. *tener* ('have')	a. *tien-e*	a. *teng-a*
	b. *decir* ('say')	b. *dic-e*	b. *dig-a*

Though there is a unified class of morphological forms in Spanish that we call subjunctive, "subjunctive does not constitute a syntactically uniform object, either cross-linguistically or even within the same language" (Quer 2006: 661). Quer (2001) and Kempchinsky (2009) have split the broad category of subjunctive mood morphology into two different subclasses, largely on the basis of the factors that trigger its presence. *Intensional subjunctive* forms are subjunctive mood forms which are triggered (obligatorily) by preceding lexical items such as *para que* ('so that'). *Polarity subjunctive* forms, on the other hand, are subjunctive forms that are selected by contextual factors such as the presence/absence of presupposition (Borgonovo, Bruhn de Garavito & Prévost 2015). For the sake of brevity, the present paper will focus exclusively on intensional subjunctive forms.

Intensional subjunctive forms, as illustrated in (1) and (2), are triggered by verbs as well as other classes of lexical items. In (1), it is the verb *quiere* ('wants') that selects for the subjunctive mood form *ganen* ('win') in the subordinate clause. In (2), on the other hand, it is the complementizer *para que* ('so that') that selects for the subjunctive mood form *dé* ('give'). In both cases, subjunctive mood selection is obligatory, as indicated by the ungrammaticality of the indicative mood forms *ganan* ('win') and *da* ('give').

(1) John quiere   que los Yanquis ganen/*ganan   esta noche.
    John want.3SG that the Yankees win.3SG.SUBJ/*IND tonight
    "John wants the Yankees to win tonight."

(2) Stan maulla para que Michael le   dé/*da   atún barato.
    Stan meows for   that Michael CL.3SG give.3SG.SUBJ/*IND tuna cheap
    "Stan meows so that Michael will give him cheap tuna."

It is not the case, however, that all verbs and complementizers select for subjunctive mood morphology, as illustrated in (3) and (4), where the verb *cree* ('believe') and the complementizer *porque* ('because') select for indicative, rather than subjunctive, morphology.

(3) John cree   que los Yanquis *ganen/ganan   esta noche.
    John believes.3SG that the Yankees win.3SG.*SUBJ/IND this night
    "John believes that the Yankees are winning tonight."

(4) Stan maulla porque Michael le   *dé/da   atún barato.
     Stan meows because Michael CL.3SG give.3SG.*SUBJ/IND tuna cheap
     "Stan meows because Michael gives him cheap tuna."

Unlike subjunctive mood morphology in polarity subjunctive structures (e.g., adjectival relative clauses: Pérez-Leroux 1998; negated epistemics: Kempchinsky 2009; or indirect commands: Perez-Cortes 2016), the lexically-selected subjunctive mood in (1) and (2) does not significantly alter the meaning of the sentences in which it appears. Consequently, if a speaker were to use non-target, indicative mood morphology after *para que*, his Spanish-speaking listeners would have no difficulty understanding his intended message, despite its ungrammaticality. It is clear, therefore, that the subjunctive mood morphology in the subordinate clauses of (1) and (2) does not *contribute* novel semantic content to the utterance as much as it *reflects* the semantic content of the preceding lexical items that selected for its presence.

   Kempchinsky (2009) offers a syntactic analysis of intensional subjunctive mood which is consistent with this proposition. For Kempchinsky, who works in a generative grammatical framework, lexical items that trigger subjunctive such as *quiere* ('want') are distinctive from other lexical items, such as *cree* ('believe'), in two critical ways. First, they are marked by a so-called W ('Worlds') feature, which signals a shift in the speaker's epistemic model, thereby introducing "a set of future worlds" that are "anchored to the matrix subject" (Kempchinsky 2009: 1797). Second, as a result of this W marking, these special lexical items only select for a certain type of sentential complement, namely, a Force phrase (ForceP)[5] that is headed by the uninterpretable Worlds feature, *u*W.[6] This feature, like other uninterpretable features in generative syntax, is checked and deleted via Agreement, in this case, agreement with the interpretable W feature represented in the complex verbal head. In summary, lexical items that select for subjunctive mood, such as *quiere* ('want'), are specially marked by the interpretable feature W, which leads them to subcategorize for a complement with the uninterpretable feature *u*W. This uninterpretable

---

5. Rizzi (1997) proposed that what was previously known as the complementizer phrase (CP) can be further split into smaller syntactic elements, including ForceP, TopicP, and FocusP. In this paper, I make reference to ForceP, rather than CP, since this is the terminology employed by Kempchinsky (2009), whose analysis I am citing in this particular section. Nonetheless, none of the argumentation presented here hinges on this subtle syntactic distinction.

6. The exact syntactic and/or semantic features which are operative in such structures differ across different theoretical frameworks. Nonetheless, as pointed out by Poplack, Lealess & Dion (2013: 158), almost all contemporary analyses of lexically-selected subjunctive mood agree on a couple of fundamental points, namely, that "subjunctive morphology is only permitted in certain complements, and that these are determined by some semantic feature(s) carried by the governing element, with which the relevant features of the complement must agree".

feature is then checked and deleted via agreement with the W feature in the verbal morphology. Lexical items that do not select for subjunctive, on the other hand, are *not* marked by the W feature, meaning that they do *not* select for complements marked with the uninterpretable *u*W feature and, consequently, do *not* require the presence of subjunctive morphology to check/delete that feature.

2.2   Acquisition of subjunctive mood by Spanish native speakers

How are intensional subjunctive forms acquired by native speakers of Spanish? To answer this question, I will briefly review previous research on the acquisition of intensional (lexically-selected) subjunctive forms by monolingual children, bilingual children, and adult HSs, respectively. Prior to doing so, however, it is important to sketch out the learning task faced by native Spanish speakers as they acquire and maintain intensional subjunctive knowledge.

If we assume Kempchinsky's (2009) analysis, native speakers learning intensional subjunctive mood in Spanish must acquire the following knowledge. First, they must acquire the W and *u*W features, respectively, since it is these abstract features that drive the presence of subjunctive mood in Spanish. Second, they must learn the lexical items (e.g., *para que*) that are marked with a W feature and, therefore, subcategorize for a ForceP headed by the *u*W feature. If a speaker acquires the relevant W feature but does not learn which lexical items are W-marked, then she may end up not producing subjunctive mood in certain expected contexts (e.g., after *para que*), even when (a) she has acquired the relevant abstract features and (b) she is capable of producing subjunctive mood in other contexts (e.g., with *quiere que*).

In addition, native speakers acquiring intensional subjunctive mood must learn how subjunctive is instantiated on different verbs in Spanish. In other words, producing target-like subjunctive mood forms requires what Lardiere (2005) has referred to as "morphological competence", e.g., knowing that the subjunctive mood form of *comer* ('eat') is *coma* and not *come*. It is possible, in theory, that a Spanish speaker could acquire the W and *u*W features, learn which lexical items are W-marked, and yet *still* not produce target-like subjunctive mood morphology on certain verbs whose subjunctive mood inflections they simply have not learned. It is apparent, therefore, that difficulties in any of these three areas could presumably lead speakers to produce or accept non-target indicative forms after lexical selectors such as *quiere* or *para que* in Spanish.

Monolingual children appear to very quickly acquire intensional subjunctive mood forms in Spanish (Blake 1983; Montrul 2004). Not only do they produce subjunctive with *para que* very early (as early as 2;6: López Ornat 1994), monolingual

children also appear to produce subjunctive with *para que* very accurately (Pérez-Leroux 1993). Nonetheless, because researchers have not systematically studied their non-target mood productions, very little is known about the steps which monolingual children take as they master lexically-selected subjunctive mood.

Like monolingual children, bilingual children appear to quickly acquire knowledge of intensional subjunctive mood forms, producing subjunctive with *para que* as early as age 3;1 (Silva-Corvalán 2014). However, despite producing intensional subjunctive forms quite early, bilingual children also have been shown to produce intermediary, non-target forms as they develop their knowledge of lexically-selected subjunctive mood in Spanish. The most striking example of such an intermediary step comes from Nico, a simultaneous Spanish-English bilingual child who produced a hybrid subjunctive/indicative verbal form (5) at age 2;11 (Silva-Corvalán 2014).

(5) para que el   pasto crezque   con la   arena.
     so   that the grass grow.SUBJ/IND with the sand.
     "So that grass may grow with the sand."

The target subjunctive form of the verb *crecer* ('grow') involves both a root change (from *cre[s]-* to *cre[sk]-*) and a thematic vowel shift (*-e* to *-a*). Interestingly, Nico produces the target subjunctive root but *not* the thematic vowel shift. On one hand, the fact that Nico produces the target subjunctive root suggests that he knows that *para que* selects for a subjunctive mood complement. On the other hand, the fact that he does *not* produce the thematic vowel shift suggests that Nico may not yet know how subjunctive is instantiated on *crecer*, e.g., that *crecer* is in the *-er* verb class. Nico's "hybrid" mood production reminds us that Spanish speakers learning subjunctive must learn not only syntactic features (and the lexical items that select for their presence) but also how those features are instantiated on different verbs in Spanish.

While a number of previous studies have examined adult HSs' knowledge of subjunctive mood in Spanish (e.g., Montrul 2009; Montrul & Perpiñán 2011), only a few previous studies isolate HSs' knowledge of intensional (as opposed to polarity) subjunctive. For the sake of brevity, I discuss only two such studies here. Van Osch and Sleeman (2018) tested HSs' oral production of subjunctive mood with *querer que*. Unlike the monolingual controls, who produced nearly 100% subjunctive with *querer que*, the HSs in the study produced subjunctive mood approximately 80% of the time, otherwise relying on non-target indicative mood forms. Perez-Cortes (2016) found similar results in a study of HSs' productive and receptive knowledge of subjunctive mood with *querer que*. In a written production task, advanced proficiency HSs produced subjunctive over 80% of the time with *querer que*. In an

acceptability task, these same HSs also correctly rejected non-target *querer que* + indicative forms over 80% of the time.

Together, these two studies show that advanced proficiency HSs are accurate (e.g. above 80%) in both productive and receptive tests of intensional subjunctive knowledge in Spanish, suggesting that they may have target-like underlying featural representations of subjunctive mood. Nonetheless, these studies also show that HSs differ from Spanish-dominant speakers in both productive and receptive tasks. Even at advanced proficiency levels, HSs still produce and accept non-target indicative mood forms in expected subjunctive contexts, raising the question of how their subjunctive knowledge differs from that of Spanish-dominant speakers. What causes HSs, even at advanced proficiency levels, to sometimes produce and/or accept non-target mood forms?

In the next two sections, I test this question by exploring the role of lexical frequency in HSs' expressed knowledge of intensional subjunctive mood with *para que* in Spanish.

3.   Experiment 1: HSs' production of intensional subjunctive mood forms

3.1   Introduction

We know, based on previous research, that HSs of Spanish produce less subjunctive in expected subjunctive contexts than Spanish-dominant speakers. However, it is unclear what factors make HSs more or less likely to produce non-target indicative in an expected subjunctive context. The goal of Experiment 1, therefore, is to shed light on this question by testing whether HSs are more likely to produce intensional subjunctive mood with frequent, rather than infrequent verbs.

Based on the findings from Dorian (1981), Gal (1989), and Montrul et al. (2014), all of whom present evidence that HSs' production of grammatical forms may be more accurate with more frequent lexical items, it is hypothesized that HSs in the present study will produce more intensional subjunctive mood with higher frequency verbs than with lower frequency verbs.

3.2   Participants

There are two experimental groups in the present study: advanced proficiency HSs (*n* = 22; henceforth, AdvHSs) and Spanish-dominant controls (*n* = 20; henceforth, SDCs). Participants in both groups were undergraduate students, graduate students or employees at a state university.

The groups differed from one another in both age of acquisition of English (AofA Eng) and language dominance. The AdvHSs, all of whom scored at least 40/50 on the DELE proficiency test (Duffield & White 1999; Bruhn de Garavito 2002; Montrul & Slabakova 2003, *inter alia*) are native speakers of Spanish who began acquiring English, in the United States, between birth and age 6. The SDCs, on the other hand, are native speakers of Spanish who began acquiring English in the United States after immigrating at age 13 or later.[7] As shown in Table 2, both groups are highly bilingual, providing relatively high self-proficiency ratings in their weaker languages. Nonetheless, the groups differed from one another in the directionality of their language dominance. While the AdvHSs' self-ratings were significantly higher in English ($M = 9.82$, $SD = 0.66$) than Spanish ($M = 8.41$, $SD = 1.37$), $t(21) = 4.830$, $p < .001$, the SDCs' self-ratings were significantly higher in Spanish ($M = 9.95$, $SD = 0.22$) than in English ($M = 7.70$, $SD = 1.34$), $t(19) = 6.957$, $p < .001$.

Table 2. Summary of participant groups in Experiment 1

Participant group	AofA English	DELE proficiency	Spanish self-rating	English self-rating
AdvHSs (*n* = 22)	Age 6 or earlier	Mean: 42.41 SD: 1.74	Mean: 8.41 SD: 1.37	Mean: 9.82 SD: 0.66
SDCs (*n* = 20)	Age 13 or later	Mean: 45.60 SD: 2.62	Mean: 9.95 SD: 0.22	Mean: 7.70 SD: 1.34

3.3   Experimental task

Participants in the study completed a Contextualized Elicited Production Task (CEPT), as well as two other experimental tasks that are not reported here. The goal of the CEPT, which was presented via PowerPoint, was to test participants' oral production of subjunctive mood forms.

First, participants read (and listened to) short contexts (in Spanish) taking place within a department store setting, where they were asked to imagine that they were buying different gifts for friends and family members. After each context, a sentence fragment appeared on screen. Participants were asked to express their wishes to a clerk at the department store by completing each sentence fragment (orally) using a form of a verb in parentheses, as well as any other words that they wished to use. A sample context and prompt, translated to English, is presented in (6).

---

7.   For more on the importance of using bilingual control groups in HS acquisition research, see Pascual y Cabo & Rothman (2012).

(6)   Context: My uncle collects a lot of books. He needs an extendable ladder to move the books on the highest shelves of his bookcase.

Prompt: You say to the clerk:

Fragment: Busco una escalera extensible para que mi tío (MOVER)…

"I am looking for an extendable ladder so that my uncle…"

Expected: *Busco una escalera extensible para que mi tío mueva*.SUBJ *los libros…*

"I am looking for an extendable ladder so that my uncle moves. SUBJ the books…"

Broadly speaking, there were two experimental conditions which are relevant for the present study.[8] The first condition is the Intensional Subjunctive condition (n = 12). In this condition, the sentence fragments included *para que* and, therefore, were expected to elicit subjunctive. The second condition was the Intensional Indicative condition (n = 6).[9] In this condition, the sentence fragments included *porque* and, therefore, were expected to elicit indicative rather than subjunctive. Participants who are sensitive to the mood selectional requirements of these two Spanish complementizers are expected to produce subjunctive with *para que* but not with *porque*.

To test the role of lexical frequency on HSs' subjunctive production, the items in the Intensional Subjunctive condition were subdivided into Higher Frequency (n = 6) and Lower Frequency (n = 6) conditions, respectively, based on Davies' (2006) lexical frequency rankings, which are calculated based on a 20-million-word corpus consisting of written (67%) and oral (33%) Spanish from both Spain (44%) and Latin America (56%) during the period of 1970 to 2000. Note that lower numbers, in this case, indicate higher frequency. An independent samples t-test revealed that the Higher Frequency verbs (M = 115.70, SD = 90.50) were significantly more frequent than Lower Frequency verbs (M = 419.67, SD = 104.71), $t(10)$ = 5.384, $p < .001$. A list of the twelve verbs used (and their frequency rankings) is presented in Table 3.

---

8.   There are two additional conditions, as well as fillers, which I do not present here.

9.   There are twice as many items in the Intensional Subjunctive condition because in the original design, this condition was actually divided into two sub-conditions. In the "baseline" sub-condition (n = 6), as exemplified in (6) above, the sentence fragments presented to participants did not include any subjunctive mood morphology. In the "prime" sub-condition (n = 6), however, the sentence fragments presented to participants included a verb with subjunctive mood morphology ("prime") *prior* to the target verb. It was hypothesized that exposure to this subjunctive mood "prime" would increase the AdvHSs' probability of producing subjunctive mood morphology. Nonetheless, the results of the data analysis revealed that the AdvHSs were no more likely to produce subjunctive mood in the presence of a prime, $p > .4$, OR = 0.76. Consequently, the prime and baseline conditions were collapsed into one condition with a n of 12.

Table 3. Verbs used in intensional subjunctive condition of CEPT

Higher frequency verbs		Lower frequency verbs	
Verb	**Davies' ranking**	**Verb**	**Davies' ranking**
hacer ('do')	25	*tocar* ('touch')	325
llegar ('arrive')	66	*correr* ('run')	332
parecer ('seem')	81	*usar* ('use')	380
salir ('leave')	111	*mover* ('move')	402
tomar ('take')	122	*observar* ('observe')	478
ganar ('win')	286	*romper* ('break')	601
Total: 6	Mean: 115.70	Total: 6	Mean: 419.67
	SD: 90.50		SD: 104.71

3.4   Results and statistical analysis

The researcher listened to participants' responses and coded them as either '1' (subjunctive) or '0' (indicative). All other response types (e.g., future tense) were excluded from the analysis. The data were analyzed using a Generalized Linear Mixed Model (GLMM) with Group (SDCs or AdvHSs) and Verb Frequency (Higher Frequency or Lower Frequency) as fixed factors and Subject and Lexical Item as random factors. Because the dependent variable is binary, the GLMM generates *predicted probabilities* (expressed as probability or odds) that participants will produce subjunctive in a given condition. In the paragraphs that follow, I present *p*-values, which evaluate differences between groups/conditions, as well as odds-ratios (OR's), which evaluate the size of those differences.[10]

As shown in Table 4, the SDCs performed exactly as expected, producing subjunctive with *para que* (*M* = 99.5%) but not with *porque* (*M* = 0.9%). Based on this finding, we can confirm the description of these forms presented in Section 2, namely, that (a) *para que* obligatorily selects for subjunctive mood and (b) *porque* obligatorily selects for indicative mood.

Table 4. Predicted probabilities of subjunctive production by group, complementizer

Group	Comple-mentizer	Exp. mood	Log odds	Standard error	Odds	Proba-bility	CI-lower	CI-upper
SDC	*para que*	Subj	5.236	0.825	187.917	99.5%	97.4%	99.9%
	porque	Ind	−4.703	0.853	0.009	0.9%	0.2%	4.6%
AdvHS	*para que*	Subj	1.866	0.422	6.462	86.6%	73.7%	93.7%
	porque	Ind	−5.958	1.100	0.003	0.3%	0.0%	2.2%

---

10. For more on the use of OR's as effect sizes, see Durlak (2009).

The AdvHSs, however, exhibited a slightly different pattern, producing both subjunctive ($M = 86.6\%$) and indicative ($M = 13.4\%$) mood morphology with *para que*. A GLMM revealed that the AdvHSs' predicted probability of producing subjunctive with *para que* was significantly lower than the predicted probability for the SDCs, $p < .001$, OR = 29.08.

Though the AdvHSs differed from the SDCs in their *quantitative* production of subjunctive, they exhibited the same *qualitative* knowledge of the mood selectional requirements of *para que* and *porque*. A within-group comparison showed that the AdvHSs' predicted probability of producing subjunctive with *para que* ($M = 86.6$) was significantly higher than their probability of producing subjunctive with *porque* ($M = 0.3$), $p < .001$, OR = 2154.0. Rather than alternating randomly between indicative and subjunctive moods, the AdvHSs show a systematic tendency to produce subjunctive with *para que* and indicative with *porque*, just like the SDCs.

Thus far, we have seen that the AdvHSs, unlike the SDCs, sometimes produce indicative mood morphology after *para que*. To explore the effect of lexical frequency on this linguistic behavior, we now turn to a second GLMM run exclusively with the *para que* items.

Table 5. Predicted probabilities of producing subjunctive with *para que* by frequency

Group	Complementizer	Verb freq.	Log odds	Standard error	Odds	Probability	CI-lower	CI-upper
SDC	*para que*	Higher	5.130	1.138	169.017	99.4%	94.8%	100.0%
		Lower	5.087	1.135	161.903	99.4%	94.6%	100.0%
AdvHS	*para que*	Higher	2.846	0.570	17.219	94.5%	84.7%	98.2%
		Lower	1.184	0.516	3.267	76.6%	53.9%	90.1%

Results of the second GLMM (Table 5) confirm that lexical frequency exerts a substantial effect on the AdvHSs' probability of producing subjunctive with *para que*. The AdvHSs' predicted probability of producing subjunctive was significantly higher with Higher Frequency verbs ($M = 94.5\%$) than with Lower Frequency verbs ($M = 76.6\%$), $p < .01$, OR = 5.27, suggesting that lexical frequency modulates their production of subjunctive mood morphology in Spanish. For the SDCs, however, lexical frequency did not affect subjunctive production, $p > .9$, OR = 1.04.

3.5   Discussion of Experiment 1

The AdvHSs in Experiment 1 appear to have clear and systematic knowledge of intensional subjunctive mood in Spanish, as indicated by (a) their high rate of subjunctive production with *para que* and (b) their very low rate of subjunctive

production with *porque*. It is reasonable to conclude, therefore, that these AdvHSs have learned the selectional requirements of the complementizer *para que*, specifically that it selects for a ForceP headed by the *u*W feature.

It is not the case, however, that the HSs perform identically to the SDCs, given that only the HSs sometimes produce non-target indicative forms after *para que*. The fact that the HSs exhibit this tendency more often with infrequent verbs, such as *observar* ('observe'), than frequent verbs, such as *llegar* ('arrive'), raises the possibility that HSs' failure to produce subjunctive mood in expected subjunctive contexts may be the result of not knowing the subjunctive instantiation of certain verbs, rather than not having acquired the relevant *u*W feature. In other words, HSs who produce these non-target forms may have acquired the underlying *u*W feature but not learned how that particular feature is instantiated on certain, less commonly used verbs in Spanish. When a HS produces a non-target form like *para que observa$_{IND}$, therefore, it may be because he simply has not learned that the subjunctive mood instantiation of the verb *observar* is *observe*$_{SUBJ}$.

4.   Experiment 2: HSs' acceptance of intensional subjunctive mood forms

4.1   Introduction

In Experiment 1, we saw that the AdvHSs differed from the SDCs in their *production* of intensional subjunctive mood forms, specifically by producing non-target indicative mood morphology after the complementizer *para que*. Based on the fact that the HSs produced more non-target indicative forms with less frequent verbs, it was argued that their non-target productions may result from not knowing how subjunctive mood features are instantiated on certain verbs. Under this explanation, HSs producing *para que observa$_{IND}$ do so because they simply do not know that the subjunctive mood instantiation of the verb *observar* is *observe.*

A second possibility, however, is that the HSs who produce non-target indicative mood after *para que* do so as a result of on-line performance effects rather than gaps in lexical knowledge. Under this explanation, HSs producing *para que observa$_{IND}$ know that the subjunctive mood instantiation of *observar* is *observe* but simply fail to produce that form in real time, perhaps due to performance pressure (e.g., Missing Surface Inflection Hypothesis: Prévost & White 2000), reduced activation (Putnam & Sánchez 2013) of *observe*, or other unknown factors associated with on-line spoken production of the HL.

In Experiment 2, I shed light on each of these two possibilities by testing a second group of HSs' *recognition*, rather than production, of subjunctive mood morphology. The logic of testing a comparable group of HSs' recognition of subjunctive

mood morphology is as follows. If HSs' divergent production of non-target indicative mood forms is attributable to gaps in lexical knowledge, then we should expect to find two specific patterns. First, the HSs should again diverge from the SDCs. Second, this divergence should be most apparent with lower frequency verbs, given that these are the verbs whose subjunctive mood instantiations HSs are least likely to know. If, on the other hand, HSs' divergent production of non-target indicative mood forms is attributable to performance effects, then we should expect a comparable group of HSs to converge with the SDCs in a receptive task of subjunctive mood knowledge, given that receptive tasks do not require on-line production and, therefore, should not lead to performance difficulties by the HSs.

4.2   Participants

The 18 AdvHSs and 15 SDCs in Experiment 2 were highly comparable to the participants in Experiment 1 with respect to DELE proficiency score as well as Spanish and English self-ratings. Because both studies were conducted at the same public university, nine participants in Experiment 2 (3 AdvHSs and 6 SDCs) also participated in Experiment 1. The inclusion criteria used for each group in Experiment 2 were identical to the criteria used in Experiment 1 with one exception. Unlike in Experiment 1, where all HSs acquired English at age 6 or earlier, in Experiment 2 there were two AdvHSs who began acquiring English at ages 7 and 8, respectively.

As in Experiment 1, the AdvHSs and SDCs, once again, were highly bilingual (e.g., rating their weaker language no less than 7.6/10), though they differed in the directionality of their language dominance. While the AdvHSs rated their English ($M = 9.44$, $SD = 0.78$) significantly higher than their Spanish ($M = 7.67$, $SD = 1.08$), $t(17) = 7.518$, $p < .001$, the SDCs rated their Spanish ($M = 9.93$, $SD = 0.26$) significantly higher than their English ($M = 7.77$, $SD = 1.05$), $t(14) = 8.765$, $p < .001$. A summary of the two groups is provided below in Table 6.

Table 6. Summary of participant groups in Experiment 2

Participant group	AofA English	DELE proficiency	Spanish self-rating	English self-rating
AdvHSs (*n* = 18)	Age 8 or earlier	Mean: 41.83 SD: 1.72	Mean: 7.67 SD: 1.08	Mean: 9.44 SD: 0.78
SDCs (*n* = 15)	Age 13 or later	Mean: 46.93 SD: 2.15	Mean: 9.93 SD: 0.26	Mean: 7.77 SD: 1.05

4.3   Experimental task

Participants in the study completed a Contextualized Acceptability Judgment Task (CAJT), as well as another experimental task not presented here. There were two goals of the CAJT, which was completed via PowerPoint. The first goal was to test whether participants recognized target and non-target mood morphology after the complementizer *para que*, specifically by accepting subjunctive mood morphology and rejecting indicative mood morphology. The second goal was to test whether participants' ability to recognize non-target mood morphology was modulated by the lexical frequency of the verb in each experimental item, in other words, whether HSs were more likely to accept non-target *para que* + indicative items with infrequent Spanish verbs.

In the CAJT, participants read short, 1–2 sentence contexts in Spanish. The last sentence of each context was a fragment ending with either *para que* or *porque*. Beneath these sentences, participants saw the remainder of the sentence fragment, which they were then asked to judge using a 1–4 Likert scale where 1 signified "sounds very odd" and 4 signified "sounds very good". A sample item, taken from the ParaQueIndicative condition is presented below in (7). In this particular experimental item, participants would be expected to select '1' or '2', given the non-target indicative mood morphology on the verb *observar* ('observe').

(7)   Context: Ana quiere ser bióloga. Su mamá le compra un microscopio para que ella…
      "Ana wants to be a biologist. Her mom buys her a microscope so that she…
      Follow-up: *observa   los microbios
                  observe.IND the microbes
                  observes microbes"

      1-------------------------------2-------------------------3-------------------------------4
      Sounds very odd   Sounds odd   Sounds good   Sounds very good

There were two relevant experimental conditions ($n = 6$ each), each of which included the complementizer *para que*. In the ParaQueSubjunctive condition, *para que* was followed by subjunctive mood, which is the target mood form. In the ParaQueIndicative condition, on the other hand, *para que* was followed by indicative mood, which is a non-target mood form. Participants who are sensitive to the mood selectional requirements of *para que* are expected to accept items in the ParaQueSubjunctive condition and reject items in the ParaQueIndicative condition.

To test the role of lexical frequency on participants' acceptance patterns, the verbs in each condition (Table 7) were divided into Higher Frequency ($n = 6$) and Lower Frequency ($n = 6$) subgroups, in accordance with Davies' (2006) frequency

rankings.[11] An independent samples *t*-test revealed that the Higher Frequency verbs ($M = 83.17$, $SD = 41.73$) were significantly more frequent than the Lower Frequency verbs ($M = 264.00$, $SD = 125.80$), $t(10) = 3.342$, $p < .01$.

Table 7. Verbs used in *para que* conditions of CAJT*

Higher frequency verbs		Lower frequency verbs	
Verb	Davies' ranking	Verb	Davies' ranking
hacer ('do')	25	*conocer* ('know')	124
ver ('see')	37	*trabajar* ('work')	183
hablar ('talk')	92	*recordar* ('remember')	215
salir ('leave')	111	*abrir* ('open')	252
volver ('return')	112	*correr* ('run')	332
tomar ('take')	122	*observar* ('observe')	478
Total: 6	Mean: 83.17	Total: 6	Mean: 264.00
	SD: 41.73		SD: 125.80

* Of the 12 verbs that appeared in the CAJT (Experiment 2), only five (41.7%) were verbs that were used in the CEPT (Experiment 1). For maximum comparability of the two tasks, it would have been ideal to select the same 12 verbs in both tasks. Nonetheless, both experiments shed light on lexical frequency effects, given that in both experiments, the verbs categorized as High Frequency are significantly more frequent than the verbs categorized as Low Frequency.

4.4   Results and statistical analysis

Participants' responses were recoded into a binary variable, such that responses of '3' or '4' were classified as '1' (accept) and responses of '1' or '2' were classified as '0' (reject).[12] Dividing the responses in this way allowed the researcher to analyze the data using binary logistic regression. The resulting binary data were analyzed using a Generalized Linear Mixed Model (GLMM) with Group (SDCs or AdvHSs),

---

11. There were two versions of the experiment. In each version, participants saw 3 Higher Frequency and 3 Lower Frequency verbs in both of the *para que* conditions. Participants in Version A, for example, saw the verb *salir* ('leave') in the ParaQueSubjunctive condition while participants in Version B saw *salir* in the ParaQueIndicative condition.

12. As a reviewer pointed out, converting 4-point scalar data to binary data results in a loss of information, namely, information about the *strength* of participants' acceptances (and rejections) of experimental items. I do not believe, however, that such information is critical to the arguments presented in this paper. If HSs, for example, are more likely than the SDCs to use '2' ratings (instead of '1' ratings) when evaluating ungrammatical sentences, what does this really indicate about their underlying knowledge of lexically-selected subjunctive mood? Given HSs' well-known tendency to "over-accept" ungrammatical sentences (Polinsky 2016), it is unclear what such a finding would illustrate.

Mood (Subjunctive or Indicative) and Verb Frequency (Higher Frequency or Lower Frequency) as fixed factors and Subject and Lexical Item as random factors.

As shown in Table 8, the SDCs performed as expected, accepting ParaQue-Subjunctive items almost categorically (93.7%) and only rarely accepting ParaQueIndicative items (4.2%). Results of the GLMM confirmed that the SDCs' odds of accepting ParaQueSubjunctive items were significantly higher than their odds of accepting ParaQueIndicative items, $p < .001$, OR = 343.78, thereby confirming the description of *para que*'s mood selection requirements presented above.

Table 8. Predicted probabilities of acceptance by group, mood

Group	Comple- mentizer	Mood	Log odds	Standard error	Odds	Proba- bility	CI-lower	CI-upper
SDC	*para que*	Subj	2.701	0.469	14.895	93.7%	85.5%	97.4%
		*Ind	−3.139	0.551	0.043	4.2%	1.4%	11.4%
AdvHS	*para que*	Subj	2.741	0.441	15.502	93.9%	84.7%	98.2%
		*Ind	−1.558	0.329	0.211	17.4%	9.8%	28.9%

The AdvHSs, however, demonstrated a slightly different pattern from the SDCs. Though the AdvHSs strongly accepted ParaQueSubjunctive items (93.9%), like the SDCs, they also showed a marginal tendency (17.4%) to accept non-target ParaQueIndicative items. Pairwise comparisons indicated that this tendency was stronger in the AdvHSs, who were significantly more accepting of non-target ParaQueIndicative items than the SDCs, $p < .05$, OR = 4.86.

The fact that the AdvHSs are more accepting of non-target items does not, however, imply that they lack sensitivity to the mood selection requirements of *para que*. Pairwise within-group comparisons revealed that the AdvHSs were significantly more accepting of grammatical ParaQueSubjunctive items than non-target ParaQueIndicative items, $p < .001$, OR = 73.63. It is clear, therefore, that the AdvHSs make the same qualitative distinctions as the SDCs.

Table 9. Predicted probabilities of accepting *para que* + indicative by group, frequency

Group	Condition	Freq.	Log odds	Standard error	Odds	Proba- bility	CI-lower	CI-upper
SDC	*para que* +Indicative	Higher	−3.143	0.768	0.043	4.1%	0.9%	16.3%
		Lower	−3.136	0.767	0.043	4.2%	1.0%	16.4%
AdvHS	*para que* +Indicative	Higher	−2.156	0.499	0.116	10.4%	4.1%	23.8%
		Lower	−0.961	0.390	0.383	27.7%	14.8%	45.7%

Is the AdvHSs' tendency to accept non-target indicative mood forms with *para que* modulated by lexical frequency, as was the case in Experiment 1? Further analyses (Table 9) revealed that the AdvHSs were marginally more likely to accept ParaQueIndicative items with Lower Frequency verbs (M = 27.7%) than with Higher Frequency verbs (M = 10.4%), p = .054, OR = 3.30, a finding that closely mirrors the lexical frequency effect found in the mood production of the HSs in Experiment 1. Between-group comparisons built on this finding by revealing that the AdvHSs were more likely than the SDCs to accept ParaQueIndicative items with Lower Frequency verbs (p < .01, OR = 8.80) but not with Higher Frequency verbs (p = .26, OR = 2.68).

4.5   Discussion of Experiment 2

The AdvHSs in Experiment 2 appear to have systematic *receptive* knowledge of intensional subjunctive mood in Spanish, as illustrated by (a) their high acceptance of subjunctive mood and (b) their almost equally low acceptance of indicative mood after the complementizer *para que*. Consequently, it seems clear that these AdvHSs have learned the mood selectional requirements of *para que*, namely that it selects for a ForceP headed by the uninterpretable *u*W feature.

Like the HSs in Experiment 1, however, the AdvHSs in Experiment 2 diverge slightly from the SDCs, this time by accepting a higher proportion of non-target indicative forms after *para que*. Interestingly, though, the AdvHSs' tendency to accept more non-target indicative forms than the SDCs was limited to the subset of experimental items which included Lower Frequency verbs, a finding that once again points to lexical, rather than morphosyntactic knowledge as a source of HS mood differences. In Experiment 1, I argued that HSs who *produce* non-target forms such as **para que observa*$_{IND}$ may do so because they do not know how subjunctive mood is instantiated on certain less frequent verbs. In Experiment 2, it may be the case that HSs who *accept* similar, non-target indicative forms do so because they do not *recognize* the mood marking of certain verbs, e.g., that *observa* is marked with indicative, rather than subjunctive, mood morphology.

It is important to conclude this section by pointing out a methodological detail with important implications. Because Experiment 2 was an off-line acceptability judgment task which did not require participants to produce oral speech in Spanish, it is less feasible to attribute the HSs' differences in acceptability with Lower Frequency verbs to performance pressure or processing limitations associated with activating lower frequency forms for speech production.

5.   Discussion and relevance of the findings

5.1   Summary of results

In the present paper, I set out to test HSs' productive and receptive knowledge of intensional subjunctive with *para que* and, in addition, whether that knowledge is shaped by lexical frequency. Experiment 1 revealed that advanced proficiency HSs have accurate productive knowledge of *para que*'s mood selectional requirements. When HSs differed from the SDCs by producing indicative mood after *para que*, it was usually with Lower Frequency verbs, pointing to a prominent role of lexical knowledge in HSs' productive variability with intensional subjunctive. Similarly, Experiment 2 showed that advanced proficiency HSs have accurate receptive knowledge of *para que*'s mood selectional requirements. When HSs differed from the SDCs by accepting indicative mood after *para que*, they did so most often with Lower Frequency verbs, suggesting that differences between the groups may be rooted in lexical, rather than morphosyntactic knowledge.

5.2   Implications for HL acquisition research

What are the implications of these findings for HL acquisition research? To answer this question, I will now return to the two approaches to HS differences presented in Section 1.

The Representational Differences Approach (e.g., Montrul 2002, 2008) interprets HS differences (e.g., producing less subjunctive mood than controls in an expected subjunctive condition) as evidence that HSs have different underlying grammatical representations, in this case, of the uW and W features in Spanish. Such a narrative, which places strong emphasis on between-group comparisons, overlooks the AdvHSs' clear and systematic accuracy with intensional subjunctive in both productive (Experiment 1) and receptive (Experiment 2) tasks. Not only do the AdvHSs in Experiment 1 produce subjunctive with *para que* over 80% of the time, the AdvHSs in Experiment 2 correctly reject non-target indicative mood after *para que* over 80% of the time. Furthermore, both the AdvHSs in Experiment 1 and the AdvHSs in Experiment 2 perform nearly identically to the SDCs with more frequent verbs in Spanish.

The results of Experiments 1 and 2 raise two challenging questions for this approach. First, if AdvHSs have distinct underlying representations of intensional subjunctive, then why do they perform so similarly to the SDCs? In other words, how do they make the same strong distinctions as the SDCs if they lack the same underlying linguistic knowledge? Second, if AdvHSs *do* have different knowledge

of intensional subjunctive, where exactly is this difference found in the HL grammar? Without a specific answer or answers to these questions, a Representational Differences approach does not help us to understand the patterns observed in the present study.

The Input Quality Approach (Pires & Rothman 2009; Pascual y Cabo & Rothman 2012) suggests that *some* HS differences (e.g., producing less subjunctive in expected subjunctive conditions) are attributable to differences in the HL input that HSs receive from first-generation immigrants. For this approach to successfully explain HS differences with a given property, therefore, it must be shown that first-generation immigrants are beginning to exhibit variability with that property, perhaps as a result of contact with the societal language.

Out of the 20 SDCs in Experiment 1, not a single participant produced even a single example of indicative mood after *para que* (226/226 subjunctive), suggesting that first-generation immigrants in the US are exceedingly unlikely to produce indicative mood after *para que* (if they ever produce it at all). Strengthening this claim is a recent study from Viner (2017), who conducted a large-scale corpus analysis and found that first-generation immigrants living in New York City produced exclusively subjunctive mood morphology on all 134 verbs following *para que*. The fact that Viner's data come from spontaneous conversation (e.g., rather than elicited production experiments) bolsters the claim that first-generation Spanish immigrants simply do not seem to produce indicative mood morphology after *para que*. Consequently, it is highly doubtful that HSs' production of non-target indicative mood forms after *para que* results from hearing such forms in the Spanish input provided to them by first-generation immigrants.

If neither representational differences nor input quality explain these HSs' differences with subjunctive mood morphology, then what does? Given that the AdvHSs performed both (a) well above chance and yet (b) less accurately with the lowest frequency verbs, it seems plausible to suggest that HS differences in producing and recognizing mood morphology (at least for higher proficiency HSs) are largely the result of gaps in lexical, rather than morphosyntactic knowledge.

As pointed out in Section 1, the present study is not the first to find that lexical knowledge conditions HSs' productive and receptive performance with morphosyntactic properties of the HL. Recall that Dorian (1981) and Gal (1989) found that HSs' production of verbal morphology in Gaelic and Hungarian, respectively, was most target-like with the most frequent verbs in the HL. Similarly, Montrul et al. (2014) reported that Spanish HSs' use and recognition of gender agreement in Spanish was significantly more accurate with canonical nouns than with non-canonical nouns, a difference that they attribute to the stronger links that HSs are able to form between the abstract gender feature and more frequently used or activated Spanish nouns.

Though each of these three studies offers support to the claim that HS differences can be rooted in lexical knowledge, none of them offers an obvious explanation as to why, or even how, lexical frequency effects emerge in HSs' knowledge of the HL. It is important, therefore, to briefly consider why exactly HSs with above-chance knowledge of a grammatical property, in this case subjunctive mood, might experience increased difficulties producing (and recognizing) that property on less frequent lexical items in the HL. It is not possible, on the basis of this data set, to offer a full and conclusive explanation. Nonetheless, I will use the following paragraphs to make a few admittedly speculative claims as to the nature of this apparent frequency pattern.

The first question to consider is why HSs either do not learn or temporarily fail to access the subjunctive mood instantiations of lower frequency verbs in Spanish. One possible explanation, though difficult to assess, is HL input quantity. It is quite safe to assume, of course, that HSs hear less frequent lexical items less frequently. It is possible, furthermore, that this reduced exposure to (or activation of; Putnam & Sánchez 2013) certain, infrequent subjunctive mood forms prevents HSs from developing a solidified association between those forms and abstract, subjunctive mood features, thus resulting in the patterns observed in the present study.

A possible weakness of this approach, however, is its implicit assumption that learning the subjunctive mood form of a given verb requires significant exposure to that specific form over time. In other words, this approach assumes that speakers must hear the subjunctive form *observe* before being able to produce it. It seems much more likely, at least in the case of Spanish-dominant speakers, that subjunctive mood competence includes not just the memorization of how subjunctive is mapped onto known verbal stems but also how to generate *novel subjunctive mood forms* (perhaps by analogy) from other, previously unknown Spanish verbs. In order to determine whether some of the HS differences observed in the present study are due to HSs' reduced ability to generate novel subjunctive mood forms, it would be necessary to conduct an additional study testing HSs' and Spanish-dominant speakers' mood inflections of nonce verbs, which, by definition, they could not have ever been exposed to in the Spanish input. If, in such a study, HSs were unable to produce mood inflections on nonce verbs, then perhaps some of their differences from the SDCs in this study are due to differences of HSs' ability to generate mood morphology.

A second question worth considering, as we attempt to understand HSs' differential performance with mood morphology, is why HSs produce non-target indicative mood forms in expected subjunctive contexts (with *para que*) but do not seem to produce non-target subjunctive mood forms in expected indicative contexts (with *porque*). One possible explanation for this asymmetrical behavior is the existence of special, default grammatical morphemes. If a HS must produce a form

of an infrequent verb after *para que*, and they do not know the subjunctive mood instantiation of that infrequent verb (as argued thus far), they are still obligated to produce *some* inflectional morphology on the verbal stem, given the fact that verb stems in Spanish are bound morphemes. When in this situation, HSs, just like L2 learners (e.g., McCarthy 2008), appear to resort to the use of underspecified, default morphemes, in this case indicative mood morphology, which has been argued to be the default mood morphology in Spanish (e.g., Bosque 2012; Quer 2001). While bilinguals frequently resort to these so-called default morphological preferences, the nature of this reliance on default forms still remains mysterious to researchers.

Putting aside the questions of why (and how) Spanish HSs exhibit differential knowledge of subjunctive mood morphology with frequent as opposed to infrequent verbs, it is clear that such a pattern is not unprecedented in language acquisition research. As pointed out by a reviewer, Poplack et al. (2013) have found that Quebec French speakers' use of subjunctive mood morphology with the lexical trigger *pour que* ('so that') appears to be strongly conditioned by lexical frequency. When *pour que* is followed by a highly frequent verb form, Quebec French speakers almost categorically produce the subjunctive mood. With lower frequency verbs, however, their use of subjunctive mood is far less consistent. This pattern closely mirrors the behavior of the HSs in the present study, and, in addition, suggests that certain grammatical forms, such as subjunctive morphology, can become "lexically-limited" in certain bilingual contexts.

Thus far, I have argued that lexical frequency can have a substantial impact on the extent to which HSs of Spanish differ from Spanish-dominant speakers. Nonetheless, I am by no means arguing that all HS differences from HL-dominant speakers are the result of gaps in lexical knowledge. A few intermediate proficiency HSs in the present study, whose results I did not include here, did not produce *any* subjunctive mood with *para que*, meaning that they almost certainly have a different underlying representation of mood as compared to the SDCs. In addition, despite the nearly categorical subjunctive mood performance of the SDCs, I am also not suggesting that input quality is irrelevant in shaping HSs' variable subjunctive mood production. It is more than plausible to believe that HSs' production of non-target indicative forms is at least partially attributable to input received from other HSs. Nonetheless, it is critical to note that this story would not explain how and why *those* HSs begin to differ from first-generation speakers in the first place.

In conclusion, it is my hope that researchers seeking to explain HS differences explore not only representation and input quality, but also lexical frequency, a promising variable that may help us better understand the mystery of how HSs can develop HL grammars that both resemble, and differ from, the grammars of the HL-dominant speakers that provide their primary input.

References

Blake, R. 1983. Mood selection among Spanish-speaking children, ages 4 to 12. *The Bilingual Review* 10: 21–32.

Borgonovo, C., Bruhn de Garavito, J. & Prévost, P. 2015. Mood selection in relative clauses: Interfaces and variability. *Studies in Second Language Acquisition* 37: 33–69. https://doi.org/10.1017/S0272263114000321

Bosque, I. 2012. Mood: Indicative vs. subjunctive. In *The Handbook of Hispanic Linguistics*, J. Hualde, A. Olarrea & E. O'Rourke (eds), 371–394. Malden, MA: Blackwell. https://doi.org/10.1002/9781118228098.ch19

Bruhn de Garavito, J. 2002. Verb raising in Spanish: A comparison of early and late bilinguals. In *Proceedings of the 26th Annual Boston University Conference on Language Development (BUCLD)*, B. Skarabela, S. Fish & A. Do (eds), 84–94. Somerville MA: Cascadilla Press.

Davies, M. 2006. *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*. New York NY: Routledge.

Dorian, N. 1981. *Language Death: The Life Cycle of a Scottish Gaelic Dialect*. Philadelphia PA: University of Pennsylvania Press. https://doi.org/10.9783/9781512815580

Duffield, N. & White, L. 1999. Assessing L2 knowledge of Spanish clitic placement: convergent methodologies. *Second Language Research* 15: 133–160. https://doi.org/10.1191/026765899668237583

Durlak, J. 2009. How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology* 34(9): 917–928. https://doi.org/10.1093/jpepsy/jsp004

Ellis, N. 2002. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24: 143–188. https://doi.org/10.1017/S0272263102002024

Gal, S. 1989. Lexical innovation and loss: The use and value of restricted Hungarian. In *Investigating Obsolescence*, N. Dorian (ed), 313–331. Cambridge: CUP. https://doi.org/10.1017/CBO9780511620997.024

Kempchinsky, P. 2009. What can the subjunctive disjoint reference effect tell us about the subjunctive? *Lingua* 119: 1788–1810. https://doi.org/10.1016/j.lingua.2008.11.009

Kupisch, T. & Rothman, J. 2018. Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism* 22(5): 564–582.

Lardiere, D. 2005. On morphological competence. In *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference (GASLA 2004)*, L. Dekydspotter, R. Sprouse & A. Liljestrand (eds), 178–192. Somerville MA: Cascadilla Press.

Leal-Méndez, T., Rothman, J. & Slabakova, R. 2015. Discourse-sensitive clitic-doubled dislocations in heritage Spanish. *Lingua* 155: 85–97. https://doi.org/10.1016/j.lingua.2014.01.002

López-Ornat, S. 1994. *La adquisición de la lengua española*. Madrid: Siglo XXI.

McCarthy, C. 2008. Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research* 24(4): 459–486. https://doi.org/10.1177/0267658308095737

Montrul, S. 2002. Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition* 5(1): 39–68. https://doi.org/10.1017/S1366728902000135

Montrul, S. 2004. *The Acquisition of Spanish: Morphosyntactic Development in Monolingual and Bilingual L1 Acquisition and Adult L2 Acquisition* [Language Acquisition and Language Disorders 37]. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.37

Montrul, S. 2008. *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor* [Studies in Bilingualism 39]. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.39

Montrul, S. 2009. Knowledge of tense-aspect and mood in Spanish heritage speakers. *International Journal of Bilingualism* 13(2): 239–269. https://doi.org/10.1177/1367006909339816

Montrul, S. 2016. *The Acquisition of Heritage Languages*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139030502

Montrul, S., Davidson, J., de la Fuente, I. & Foote, R. 2014. Early language experience facilitates gender agreement processing in Spanish heritage speakers. *Bilingualism: Language and Cognition* 17(1): 118–138. https://doi.org/10.1017/S1366728913000114

Montrul, S. & Perpiñán, S. 2011. Assessing differences and similarities between instructed heritage language learners and L2 learners in their knowledge of Spanish tense-aspect and mood (TAM) morphology. *Heritage Language Journal* 8(1): 90–133.

Montrul, S. & Sánchez-Walker, N. 2013. Differential object marking in child and adult Spanish heritage speakers. *Language Acquisition* 20(2): 109–132. https://doi.org/10.1080/10489223.2013.766741

Montrul, S. & Slabakova, R. 2003. Competence similarities between native and near-native speakers: An investigation of the preterite/imperfect contrast in Spanish. *Studies in Second Language Acquisition* 25: 351–398. https://doi.org/10.1017/S0272263103000159

van Osch, B. & Sleeman, P. 2018. Spanish heritage speakers in the Netherlands: Linguistic patterns in the judgment and production of mood. *International Journal of Bilingualism* 22(5): 513–529. https://doi.org/10.1177/1367006916654365

Pascual y Cabo, D. & Gómez Soler, I. 2015. Preposition stranding in Spanish as a heritage language. *Heritage Language Journal* 12(2): 186–209.

Pascual y Cabo, D. & Rothman, J. 2012. The (Il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied Linguistics* 33(4): 450–455. https://doi.org/10.1093/applin/ams037

Perez-Cortes, S. 2016. Acquiring Obligatory and Variable Mood Selection: Spanish Heritage Speakers' and L2 Learners' Performance in Desideratives and Reported Speech Contexts. PhD dissertation, Rutgers University.

Pérez-Leroux, A. T. 1993. Empty Categories and the Acquisition of wh- Movement. PhD dissertation, University of Massachusetts, Amherst.

Pérez-Leroux, A. T. 1998. The acquisition of mood selection in Spanish relative clauses. *Journal of Child Language* 25(3): 585–604. https://doi.org/10.1017/S0305000998003614

Pires, A. & Rothman, J. 2009. Disentangling sources of incomplete acquisition: An explanation for competence divergence across heritage grammars. *International Journal of Bilingualism* 13(2): 211–238. https://doi.org/10.1177/1367006909339806

Polinsky, M. 2006. Incomplete acquisition: American Russian. *Journal of Slavic Linguistics* 14(2): 191–262.

Polinsky, M. 2016. Looking ahead. In *Advances in Spanish as a Heritage Language* [Studies in Bilingualism 49], D. Pascual y Cabo (ed.), 325–345. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.49.17pol

Poplack, S., Lealess, A. & Dion, N. 2013. The evolving grammar of the French subjunctive. *Probus* 25(1): 139–195.

Prévost, P. & White, L. 2000. Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research* 16(2): 103–133. https://doi.org/10.1191/026765800677556046

Putnam, M. & Sánchez, L. 2013. What's so incomplete about incomplete acquisition? A prolegemenon to modeling heritage language grammars. *Linguistic Approaches to Bilingualism* 3(4): 476–506. https://doi.org/10.1075/lab.3.4.04put

Quer, J. 2001. Interpreting mood. *Probus* 13: 81–111. https://doi.org/10.1515/prbs.13.1.81

Quer, J. 2006. Subjunctives. In *The Blackwell Companion to Syntax*, Vol. IV, M. Everaert & H. van Riemsdijk (eds), 660–684. Malden MA: Blackwell. https://doi.org/10.1002/9780470996591.ch68

Rizzi, L. 1997. The fine structure of the left periphery. In *Elements of Grammar*, L. Haegeman (ed.), 281–337. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-011-5420-8_7

Sherkina-Lieber, M. 2015. Tense, aspect, and agreement in heritage Labrador Inuttitut: Do receptive bilinguals understand functional morphology? *Linguistic Approaches to Bilingualism* 5(1): 30–61. https://doi.org/10.1075/lab.5.1.02she

Silva-Corvalán, C. 2014. *Bilingual Language Acquisition: Spanish and English in the First Six Years*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139162531

Viner, K. 2017. Second generation NYC bilinguals' use of the Spanish subjunctive in obligatory contexts. *Spanish in Context* 13(3): 343–370. https://doi.org/10.1075/sic.13.3.02vin

Word order variation in heritage languages

Subject shift and object shift in Norwegian

Merete Anderssen[i] and Marit Westergaard[i,ii]
[i]The Arctic University of Norway / [ii]Norwegian University of Science and Technology

This study investigates two word order phenomena in Norwegian heritage language spoken in the US, subject shift (SS) and object shift (OS). SS and OS occur in syntactic environments where (pronominal) subjects and objects may either precede or follow negation. This paper explores to what extent these two phenomena in Heritage Norwegian are affected by the factors frequency and structural similarity/difference. As subjects are frequently shifted, while objects are not, SS is expected to be robust and OS vulnerable. There is generally no structural overlap between English and Norwegian in these cases; thus, cross-linguistic similarity or difference should not play a role, except in one context: questions with auxiliaries or *be*, in which the two languages allow both orders (*is he not/isn't he*), but have opposite preferences. The results show that OS is somewhat vulnerable, but SS is also affected, in that both proficient and less proficient speakers seem to overuse the word order preferred in English. We thus speculate that all heritage speakers may be affected by cross-linguistic influence in situations with complete structural overlap.

Keywords: incomplete/differential acquisition, attrition, frequency, complexity, structural overlap, cross-linguistic influence, cross-linguistic overcorrection, syntactic movement, verb second

1. Introduction

This study investigates two word order phenomena in Norwegian, subject shift (SS) and object shift (OS) in the speech of Norwegian heritage speakers in the United States. The two structures are illustrated in (1) and (2) respectively and refer to the position of pronominal subjects and objects in front of negation and other adverbs. Since non-pronominal elements typically appear following negation, it is assumed that pronominal elements are moved (or shifted) across negation (more details in Section 2).

(1) *Det liker **hun ikke**.*   (SS)
     that likes she   not
     "She doesn't like that."

(2) *Hun liker **det ikke**.*   (OS)
     she   likes it   not
     "She doesn't like it."

The last 20 years has seen the growth of a considerable body of work on heritage languages. Generally, these languages are described as having "'simplified', non-standard characteristics" (Scontras, Fuchs & Polinsky 2015: 3), with all parts of language potentially affected (see e.g. Montrul 2014 and references therein). However, the level of proficiency of these speakers varies greatly, and many reach a high competence in the heritage language (see e.g. Kupisch 2013). A recurring question within research on heritage languages relates to what causes these differences. One explanation links the lack of proficiency to a disruption in the acquisition process, due to diminishing input in the heritage language as the child gets older and more dominant in the majority language. The result is what is often referred to as 'incomplete acquisition' (see e.g. Polinsky 2006; Montrul 2002, 2008), although a more appropriate term may be 'differential acquisition' (see Kupisch & Rothman 2018). Another view is that divergent grammars in heritage speakers are the result of language attrition; the language has been completely acquired but has subsequently undergone erosion due to a lack of exposure and use (see e.g. Köpke & Schmid 2004). Whether incomplete/differential acquisition or attrition is the more likely explanation for divergent features in the heritage grammar depends to a large extent on the age at which the relevant feature is acquired and at what age the dominance shift occurs. Incomplete/differential acquisition is more likely with phenomena that are acquired relatively late, at least if the shift from the minority to the majority language occurs after school age. Of the two phenomena that we investigate here, one of them, SS, is typically acquired by the age of three (Anderssen, Bentzen, Rodina & Westergaard 2010), while the other, OS, is not fully in place until early school age (Anderssen, Bentzen & Rodina 2012).

 Putnam and Sánchez (2013) propose an approach to heritage languages in which these grammars gradually undergo reanalysis and restructuring because of influence from the majority language.[1] The extent to which the majority language exerts this influence on the heritage language depends on the degree to which the

---

[1] One important question is whether the discrepancy between heritage and non-heritage speakers is due to processing difficulties or representational differences. The approach proposed by Putnam and Sánchez (2013) assumes that heritage speakers ultimately end up with divergent representations, but processing (for production and comprehension) also plays a prominent role in the model.

heritage language is activated: the less activation, the more susceptible the heritage language will be to influence. Putnam and Sánchez explicitly state that it is activation and not frequency that determines to what extent this takes place. However, it is reasonable to assume that frequency of exposure and use will affect the level of activation in these speakers. According to this view, then, heritage speakers will be at various stages on a sliding spectrum, where decreasing activation of the heritage language causes increasing cross-linguistic influence from the majority language, resulting in restructuring of the heritage grammar. In this paper, we assume that the changes in the heritage language come about along these lines, and we follow Müller and Hulk's (2001) assumption that in order for cross-linguistic influence to take place in the grammar, there has to be superficial structural similarity between the two languages. However, we also show that the notion of structural similarity needs to be specified more carefully, as there are several ways in which languages may overlap. There might be just one overlapping structure in the two languages, or it might be that one language allows two word order alternatives (for example VO/OV), while the other only allows one (VO). In such a situation, speakers may be affected by both the similar, overlapping structure and the one that is different in the two languages. A third possibility is that both languages allow the same two word order alternatives, but have different conditions on when they are used. As structural similarity may occur in many different constellations, it is to be expected that it can have different linguistic outcomes as well.

Against this backdrop and based on previous research on SS and OS in first language acquisition (e.g. Anderssen et al. 2010, 2012) we ask the following main research question: How will the factors frequency and structural similarity vs. structural difference between the heritage and majority languages play out in Heritage Norwegian (HN) with regard to SS and OS? Our previous research on the population of HN speakers on morphosyntactic phenomena where there is (partial) structural overlap has suggested that frequency and structural difference are important factors, at least for more proficient speakers, while for less proficient speakers, structural similarity plays a more important role (e.g. Westergaard & Anderssen 2015; Anderssen, Lundquist & Westergaard 2018). As there is generally no structural overlap between Norwegian SS and OS and the equivalent structures in English, cross-linguistic similarities/differences are not expected to have an effect, but the two phenomena are used with very different frequencies, and consequently, we expect OS to be more vulnerable than SS. Yet, our results show that the heritage speakers perform in a similar way with the two structures, suggesting that frequency does not play a major role in the maintenance of these structures. We also find that structural similarity seems to have an effect also on more proficient speakers. This is found in a situation where both Norwegian and English have two possible word orders and thus complete structural overlap (i.e. for SS in questions).

2. Background

2.1 The structures

Norwegian is a verb-second (V2) language with a general requirement that the finite verb move to second position in all declaratives and most questions.[2] This is illustrated for a non-subject initial declarative in (3) and a *wh*-question in (4); in both cases the finite verb appears in a position preceding the subject.

(3) *I går*   **leste hun** *to*   *aviser.*
     yesterday read she two newspapers
     "Yesterday she read two newspapers."

(4) *Hvorfor* **leste hun** *to*   *aviser?*
     why   read she two newspapers
     "Why did she read two newspapers?"

Both SS and OS are in certain ways related to V2 syntax, since the context for the shift is dependent on the verb having moved out of the VP (that is, only then will the subject or object be adjacent to adverbs/negation). This is important for our purposes, as this means that there is normally no superficial structural similarity between the two languages in these structures, as English does not generally display V2 word order (cf. Section 2.2).

In main clause SS constructions, informationally given subjects generally occur in front of negation and new or focused subjects after (typically realized as pronouns and DPs respectively), as illustrated in (5)–(6). In embedded clauses, on the other hand, both pronouns and DPs mainly appear in the shifted position; see (7). For more detailed information on subject positions, see e.g. Mohr (2005); Cardinaletti (2004) and van Kemenade & Los (2006); specifically about Norwegian, see e.g. Nilsen (1997); Svenonius (2002); Bentzen (2009) and Westergaard (2011).

(5) *I går*   *spiste {han} ikke {?han} middag.*
     yesterday ate   he   not   he   dinner
     "Yesterday he didn't have dinner."

(6) *I går*   *spiste {?Jon} ikke {Jon} middag.*
     yesterday ate   John   not John dinner
     "Yesterday John didn't have dinner."

---

2. Non-V2 is a widespread phenomenon in wh-questions in many Norwegian dialects; see e.g. Westergaard (2009a); Westergaard, Vangsnes & Lohndal (2017).

(7) *Jeg visste at*   *{Jon/han} ikke {?Jon/?han} hadde spist middag.*
    I   knew that John/he   not   John/he   had   eaten dinner
    "I knew that Jon/he had not eaten dinner."

In corpora of everyday conversational speech, the percentage of shifted pronominal subjects has been found to be around 85–90%, in both main and embedded clauses. More specifically, Westergaard (2011) finds that the distribution of shifted vs. non-shifted pronominal subjects in a corpus of child-directed speech (Anderssen 2006) was 87.9% (1185/1348) in main and 90.1% (246/273) in embedded clauses (data from 8 adults). Furthermore, the distribution in the NoTa corpus (166 Oslo speakers) is very similar: Table 1 (adapted from Westergaard 2011) shows that pronominal subjects are shifted 84.7% (1839/2170) in main clauses and 88.2% (531/602) in embedded clauses.

Table 1. Full DP and pronominal subjects with Neg-S and S-Neg word order in main and embedded clauses in the NoTa corpus (166 Oslo speakers)

Clause type	Pronominal subject		DP subject	
	S-Neg	Neg-S	S-Neg	Neg-S
Main clauses	**84.7%** (1839/2170)	15.3% (331/2170)	3.4% (1/29)	**96.6%** (28/29)
Embedded clauses	**88.2%** (531/602)	11.8% (71/602)	**73.7%** (28/38)	26.3% (10/38)

Table 1 also shows that DP subjects are quite infrequent, making up only 1.3% (29/2199) and 5.9% (38/640) in main and embedded clauses respectively. Furthermore, DPs also behave differently from pronominal subjects in main clauses, in that they typically do not shift (only 3.4%). In embedded clauses, on the other hand, both pronominal and DP subjects are typically shifted (88.2% and 73.7%).

OS is found in contexts in which the finite verb moves out of the VP, which means that it is restricted to main clauses with only one verb.[3] DP objects must always appear after negation, and are thus ungrammatical in the shifted position, as illustrated in (8). OS has been extensively studied from various perspectives, and a number of different accounts have been proposed. For more information about OS, see e.g. Holmberg (1986, 1999); Jayaseelan (2001); Erteschik-Shir (2005); Vikner (2006); Vogel (2006); Andréasson (2008, 2010); Josefsson (2010) and Mikkelsen (2011), among many others. Pronominal objects with nominal antecedents typically shift and thus precede negation, unless they carry contrastive stress, in which case they cannot shift, as shown in (9). This means that OS is also to some extent

---

3. If there is more than one verb in the VP, the auxiliary will move to second position and the main verb will stay in a position between negation and the object, thus blocking OS, e.g. *Hun har **ikke sett ham*** "She has not seen him."

dependent on information structural factors (Anderssen & Bentzen 2012). There are further restrictions on OS in that there are a number of pronouns that do not shift, e.g. indefinites or possessives. Most importantly (since they are so frequent in the input), pronominal objects with non-nominal antecedents usually do not shift, as shown in (10) (Andréasson 2008, 2010; Anderssen & Bentzen 2012).

(8) *Peter så   {*bilen} ikke {bilen}.*
    Peter saw car.DEF not car.DEF
    "Peter didn't see the car."

(9) *Peter så   {den} ikke {DEN}.*
    Peter saw it   not THAT
    "Peter didn't see it/THAT."

(10) *Marit synes den er fin, men Peter synes **ikke det.***   (**det** = 'that it is nice')
     Marit thinks it   is nice but Peter thinks not   it
     "Marit thinks it is nice, but Peter doesn't think so."

In corpora of spontaneous everyday speech, pronominal objects with nominal antecedents shift at approximately 85%. Bentzen, Anderssen and Waldmann (2013) studied two corpora of child-directed speech (Simonsen 1990; Anderssen 2006) and show that pronominal objects with nominal and non-nominal antecedents shift at very different rates, 87% (41/47) vs. 5% (12/237). As object pronouns with non-nominal antecedents are much more frequent than pronouns with nominal antecedents, and as objects are often realized as DPs and DPs do not move, the vast majority of objects do not shift. The results from Bentzen et al. (2013) are summarized in Table 2.

Table 2. Pronominal objects with nominal and non-nominal antecedents with O-Neg and Neg-O word order in two child language corpora (80,000 adult utterances)

Type of antecedent	O-Neg	Neg-O
Nominal antecedent	**87%** (41/47)	13% (6/47)
Non-nominal antecedent	5% (12/327)	**95%** (315/327)

This corresponds closely to the results from the Nordic Dialect Corpus, where 87.6% (149/170) of object pronouns with non-nominal antecedents shift (Bentzen 2014). Furthermore, in a 13-hour sample of child-directed speech investigated in Anderssen et al. (2010), there were 157 examples of SS and only 3 examples of OS.

Summing up, contexts for SS occur much more frequently than contexts for OS in natural speech. Furthermore, while the majority of subjects are pronominal and occur in the shifted position, the majority of objects are DPs or pronouns with non-nominal antecedents and remain in situ.

2.2   Structural similarity/difference – Norwegian vs. English

As mentioned in the introduction, superficial structural similarity is taken to be a prerequisite for cross-linguistic influence (Müller & Hulk 2001), and, as discussed in the previous section, both SS and OS structures are related to verb movement and V2 word order. Since English does not generally display V2, this means that there should be no structural overlap between SS and OS in Norwegian and the corresponding English structures, since in both cases, there will be an auxiliary or a verb intervening between the subject/object and negation in main clauses in English. This is illustrated schematically in (11) for SS and (13) for OS, and with relevant examples in (12) and (14):

(11)   SS: found in non-subject-initial declarative main clauses
    a. Norwegian: *XP V S$_{pro}$ neg + XP V neg S$_{DP}$*
    b. English: *XP   S$_{pro/DP}$ **aux** neg …*

(12)   a. Norwegian: *I går   leste **han ikke** avisa*
                 yesterday read he   not paper.DEF
    b. English: *Yesterday **he DID not** read the newspaper*

(13)   OS: found in sentences with verb movement and only one verb
    a. Norwegian: *S V   O$_{pro}$ neg   + S V neg O$_{DP}$*
    b. English: *S aux   neg V O$_{pro/DP}$*

(14)   a. Norwegian: *Han leste **den ikke***
                 he   read it   not
    b. English: *he   DID   **not READ it***

However, there is one case where there is structural overlap between English and Norwegian, viz. in questions with auxiliaries or *be*, which require verb movement also in English, i.e. subject-auxiliary inversion. This is illustrated in (15a–d). Note that English displays the same word order variation that is found in Norwegian SS constructions, in that it allows both S-Neg and Neg-S.

(15)   Questions: Norwegian V2 / English S-Aux inversion
    a. *Hva kunne **ikke han** gjøre? /Hva kunne **han ikke** gjøre?*
    b. What could**n't he** do? / What could **he not** do?
    c. *Er **ikke hun** flink? / Er **hun ikke** flink?*
    d. Is**n't she** clever? / Is **she not** clever?

This means that there is total structural overlap between Norwegian and English with respect to SS in questions, as illustrated in (16). However, the preferences are different for the two languages: While S-Neg is clearly the preferred option with pronominal subjects in Norwegian (cf. the previous section), it is the opposite word

order that is preferred in English, due to the propensity for negation to attach to the auxiliary as a clitic. This means that S-Neg is a marked and infrequent word order in English, while Neg-S is marked and infrequent in Norwegian (unmarked/ frequent word order in bold here).

(16)   a. Norwegian: *(Q) V S_{pro} **neg** / (Q) V neg S_{DP}*
       b. English: *(Q) aux S neg V / **(Q) aux neg S V***

2.3   SS and OS in L1 acquisition

Norwegian children have been found to exhibit a delay in the acquisition of SS and OS, both in corpora and in experimental studies (Westergaard 2008, 2011; Anderssen et al. 2010, 2012). For SS, children seem to have a preference for the unshifted position early on, illustrated in (17), although the distribution of shifted pronominal subjects reaches adult-like levels already by age 2;6–3, and somewhat later in embedded clauses (Westergaard 2011).

(17)   *no   kan **ikke han** sove mer.*   (Ann 2;3.9)
       now can not he   sleep more
       "Now he can't sleep anymore."   Target: *No kan han ikke sove mer.*

Compared to SS, OS is more severely delayed in child language, often until the age of 6–7 (Anderssen, Bentzen & Rodina 2012). This is illustrated in (18).

(18)   *Han erta   ikke **ho**.*   (B04 5;5.17)
       he   teased not her
       "He didn't tease her."   Target: *Han erta **ho** ikke.*

It has been argued that this delay in child language is due to economy, i.e. children's tendency to avoid complexity, such as syntactic movement, something that has been found for many other movement constructions in child language (see e.g. Westergaard 2009b, 2014; Anderssen, Rodina, Mykhaylyk & Fikkert 2014).[4] Overgeneralization of movement is virtually never attested (see e.g. Radford 1992; Roeper 1999). This has been considered to be part of the conservative nature of L1 acquisition, based on findings that young children typically make errors of omission

---

4.   This view of economy is generally found within generative approaches to language (acquisition) where it is assumed that arguments with the same thematic role are base-generated in the same position, and any variation in surface order is taken to be due to syntactic movement. Thus, the alternation between *Han så **ikke** jenta* ('he saw not the girl') and *Han så **henne** ikke* ('he saw her not') is the result of the pronoun *henne* ('her') shifting past the negation in the second sentence. This movement is seen as being less economical or more complex than the alternative.

and hardly ever errors of commission (Snyder 2007). The more extended delay of OS compared to SS has been related to the very low frequency of OS in the input (cf. Section 2.1) as well as the complexity of OS, generally related to the many restrictions on the kinds of pronominal objects that undergo movement (Anderssen, Bentzen & Rodina 2012). Thus, the factors economy and complexity have been argued to be important in child language, while frequency only plays a role in connection with other factors (see e.g. Roeper 2007; Westergaard & Bentzen 2007).

2.4   Previous research on heritage Norwegian

The current paper investigates a corpus of spontaneous speech produced by 50 speakers of Norwegian heritage language in the US (more information in Section 4) that has also been studied in Westergaard & Anderssen (2015) and Anderssen et al. (2018) on other phenomena that display variation, i.e. possessives and modified definites. We briefly summarize the results of these studies here, as they are relevant for the heritage speakers' behavior with SS and OS and also form the basis of our suggested analysis.

Possessive structures in Norwegian may be pre- or postnominal (19), and it is argued that the postnominal structure is derived from the prenominal one, thus being more complex (Anderssen & Westergaard 2010; Lødrup 2011; Westergaard & Anderssen 2015). Despite the former being considerably more frequent than the latter in the input (75% vs. 25%), mono- and bilingual children have been shown to overuse prenominal possessives at an early stage (Anderssen & Westergaard 2010; Westergaard & Anderssen 2015). The Norwegian HSs, on the other hand, exhibit almost exclusive use of the postnominal possessive.

(19)   a.   *min bil*
        my   car
    b.   *bilen   min*
        car.DEF my

Westergaard & Anderssen (2015) observe that a small subset of the HSs exhibits a preference for prenominal possessives, and Anderssen et al. (2018) suggest that, while this minority is affected by cross-linguistic influence (CLI) and overuse the structures that are *similar* to English, the majority could be described as influenced by what Kupisch (2014) refers to as *cross-linguistic overcorrection* (CLO), i.e. a preference for the structure that is *different* from the dominant language. Based on these observations, Anderssen et al. (2018) divide the speakers into two groups, excluding 22 speakers who produce too few relevant structures. This leaves 28 speakers, one group that is affected by CLI (7 speakers) and another that is affected by CLO (21 speakers), referred to as the *English* group and the *Norwegian* group respectively.

Anderssen et al. (2018) also investigate double definiteness, illustrated in (20b), which is required in modified definite DPs.

(20)   a.   *en bil – bilen*
              a   car – car.DEF
         b.   *den store bilen*
              the big   car.DEF

With respect to complexity, frequency, and structural similarity vs. difference, modified definites are similar to possessives. The suffixal article is clearly *structurally different* from English and also very frequent, being used in both modified and unmodified definites. The prenominal determiner, on the other hand, is *structurally similar* to English and it is also infrequent. The results show that the heritage speakers produce two types of non-target-like structures, dropping the suffix or dropping the prenominal determiner. Furthermore, the speakers in the *English* group typically drop the suffixal article, while the *Norwegian* group is significantly more likely to drop the prenominal determiner. With respect to proficiency, using the total number of errors in modified definites as well as gender-marking (investigated in Lohndal & Westergaard 2016) as measures, the *Norwegian* group is more target-like than the *English* group.

For the majority of HSs, the results from possessives and modified definites thus suggest that frequency and CLO have a large impact. These speakers exhibit a preference for postnominal possessives and modified definites with the suffixal article as the only exponent of definiteness, that is, the alternatives that are *different* from English (CLO). However, postnominal possessives and modified definites with the suffixal article are also the more frequent structures in Norwegian. Consequently, it is not possible to determine which factor is more important, frequency or structural difference, or indeed, whether both factors are at play. This is why considering SS and OS might prove fruitful, as these structures do not involve superficial structural similarity, and CLO should therefore be irrelevant.

In both possessives and modified definites, there is partial overlap between Norwegian and English, as English has one option and Norwegian has two. When the heritage speakers use Norwegian, their dominant language (English) is also activated. For the speakers with a high proficiency in Norwegian, the inhibition of the English structures may also inhibit the similar structure in Norwegian and thus reinforce the structure that is *different* from English, while for the less proficient speakers it activates the English alternative (see the discussion section for a more articulated proposal). This is illustrated schematically for possessives and modified definites in Tables 3–4, where the shaded areas show which structures overlap and have to be inhibited in English and the resulting preference in HN is shown in italics (boldface indicates language dominance).

Table 3. CLI from English into HN in structures with partial overlap (English group)

Language\Structure	Possessives		Double definiteness	
Norwegian	*POSS N*	N POSS	*Determiner*	Suffix
English	**POSS N**		**Determiner**	

Table 4. CLO from English into HN in structures with partial overlap (Norwegian group)

Language\Structure	Possessives		Double definiteness	
Norwegian	POSS N	*N POSS*	Determiner	*Suffix*
English	**POSS N**		**Determiner**	

3.   Research questions and predictions

Since complexity (in terms of syntactic movement) has not been found to be an issue for HN speakers, we do not in principle expect it to be difficult for these speakers to produce subjects and objects in shifted positions. However, we do expect frequency to play a role. The question is whether we should consider overall frequency or frequency within a local domain. If the former, then we would expect OS to be vulnerable and SS to be unproblematic, since the latter construction is massively more frequent than the former (cf. Section 2.1). Likewise, if SS is affected at all, we would expect embedded clauses to be more problematic than main clauses, given that they are much less frequent in everyday language use (640 to 2839 in the NoTa corpus; cf. above). However, as it has been shown in a number of studies that learners are able to make fine distinctions between syntactic constructions (e.g. the micro-cue model: Westergaard 2009b, 2014), frequency should play a role in more local domains, i.e. in individual clause or construction types separately. Thus, SS should be relatively unproblematic, since most subjects appear in shifted position (2392 shifted vs. 440 unshifted subjects in the NoTa corpus), while the opposite situation holds for OS, since all DP objects and non-nominal pronouns do not shift.

   With respect to cross-linguistic influence, this has been argued to play a role when there is partial structural overlap between the two languages involved (Müller & Hulk 2001), as with possessives and double definiteness in HN. With SS and OS, on the other hand, there is generally no structural overlap, since both constructions require verb movement in Norwegian, and we consequently would not expect cross-linguistic influence to have any effect. The only exception to this is questions, where English does have verb movement, and where both word orders are grammatical (cf. Section 2.2). Thus, there is total structural overlap in this context, with

opposite preferences in the two languages. Based on previous findings, we would therefore expect groups of HSs to be affected differently: Proficient speakers should be affected by CLO and less proficient speakers by CLI. This means that speakers in the *Norwegian* group should prefer the shifted position (S-Neg), while less proficient speakers (the *English* group) should prefer the non-shifted position (Neg-S).

Summarizing, we make the following predictions for SS and OS in HN:

(21) If frequency plays a role (in local domains):
    a. OS should be vulnerable (non-shifted position more frequent than shifted position)
    b. SS should be unaffected in main and embedded clauses (shifted position more frequent than unshifted)

(22) If cross-linguistic similarities/differences play a role in contexts with total overlap:
    a. Proficient speakers should prefer SS (affected by CLO)
    b. Less proficient (attrited) speakers should prefer non-shifted order (CLI)

4. Participants

The data for the present study are taken from the Corpus of American-Norwegian Speech (CANS) (Johannessen 2015), collected through the project NorAmDiaSyn. The database currently consists of transcribed interviews/conversations with 50 Norwegian HSs, with approximately 1–2 hours of speech recorded per speaker. They learned Norwegian from birth from their parents and grandparents and English either from around age 5–6 when starting school or somewhat before their school years. This means that they are typical HSs who have experienced a shift in language dominance. However, these heritage speakers are unusual from the point of view that they are 2nd–4th generation immigrants, and thus up to several generations removed from the non-heritage variety of the language. They are also different from the HSs usually investigated because of their age; most of the speakers in CANS are approximately 70–90 years old. They are all clearly English dominant and use Norwegian only for special occasions and with very few other speakers. The majority of them are also not (or minimally) literate in Norwegian.

The first Norwegian immigrants arrived in the US in 1825 (Haugen 1953; Lovoll 1999). Even though the new arrivals came from many different dialect groups, the dialect spoken by the largest number of immigrants (from rural eastern Norway) has become dominant and is typically the only one that survives among Norwegian HSs today (Johannessen & Salmons 2015; Johannessen & Laake 2012, 2017). Johannessen and Laake (2012) propose that this variety formed the basis of

a koiné, and thus, the variety spoken by Norwegian immigrants in the US should be regarded as one lexically defined dialect (Johannessen & Laake 2017). Nevertheless, when studying any linguistic phenomenon in HN, we do so without any definitive knowledge about what the early input to these speakers was. Because of this, using homeland Norwegian as a reference point might not be appropriate. Furthermore, when it comes to SS and OS, there is some dialectal variation in Norway today, and it is relevant to consider whether these phenomena are present in the dialects spoken in the areas where most HSs came from (rural eastern Norway). According to Venås (1971), who investigated dialects in all parts of Norway except the north, SS is used throughout, except in the county of Trøndelag and in the northwest. This suggests that the relevant dialects do have SS, at least today. For OS, Bentzen (2014) shows that examples where pronouns with nominal antecedents are not shifted in The Nordic Dialect Corpus cluster around the western part of the country (the counties of Hordaland, Møre & Romsdal and Sogn & Fjordane) and Trøndelag, suggesting that rural eastern Norwegian dialects also display OS today. However, this does not guarantee that this was the case 170 years ago, nor does this mean that SS and OS were as frequent in the input in HN as in the corpora reported on in Section 2.1. Nevertheless, this is the best source that we have for a baseline, and we therefore take this as our point of departure.

5.   Results

5.1   Subject shift in heritage Norwegian

The results of our investigation of SS in the CANS corpus are provided in Table 5. First and foremost, we observe that, considering the size of the corpus, the data are relatively sparse, with only 181 contexts for SS altogether (non-subject-initial main clause declaratives and questions as well as embedded clauses). There are presumably several reasons for this: (i) The interview situation has not been conducive to eliciting many questions from the HSs, (ii) these HSs do not use Norwegian much and may therefore resort to simpler structures; thus they produce few embedded clauses, and (iii) while non-subject-initial declaratives are quite frequent in Norwegian and other V2 languages (30–40%), this is not the case in English, where subjects are favored in initial position, and the HSs may be affected by English in this respect. In fact, investigating the current speakers in the CANS corpus, Westergaard and Lohndal (2019) have attested a significantly lower production of non-subject-initial declaratives in HN than in non-HN. This means that the contexts for SS are to some extent avoided in these HSs.

Table 5. Word order in SS contexts with pronominal subjects in CANS ($n = 50$)

Clause type	S-Neg	Neg-S
Main clauses	60.0% (87/145)	40.0% (58/145)
Embedded clauses	86.1% (31/36)	13.9% (5/36)

As shown in Table 5, the HSs behave like Norwegian non-heritage speakers with regard to SS in embedded clauses, shifting pronominal subjects 86.1% (compared to 88.2% and 90.1% in the two Norwegian corpora mentioned above). An example of this is provided in (23).

(23) *det   er mye [**vi ikke** veit   på engelsk au]*
      there is a lot we not know in English also
      "There is a lot we don't know in English too."

It should be noted that the majority of the relevant embedded clauses (i.e. embedded clauses including negation) are produced by the 21 speakers in the *Norwegian* group (55.6%, 20/36), and not a single one is produced by the *English* group, suggesting that the speakers with a lower proficiency are not producing very many embedded structures at all.

In main clauses, on the other hand, the proportion of shifted subjects is much lower, only 60.0% (87/145). This is significantly different from the two non-HN corpora (cf. Section 2.1), where pronominal subjects were shifted 84.7% and 87.9% ($p < 0.001$). An example of the unshifted word order is provided in (24).

(24) *nei jeg veit   da   vi begynte på skolen så [#] kunne **ikke vi** [#] snakke*
      no I   know when we started at school so could   not we speak
      engelsk at all
      English at all
      "No, I know that when we started school, we couldn't speak English at all."

With respect to the two groups of HSs, we find only a negligible difference between them, in that the 21 speakers in the *Norwegian* group shift subjects 65.3% (49/75), while the 7 speakers in the *English* group shift slightly less, 61.1% (11/18). Note that the raw numbers are quite low here, due to the fact that so many speakers were excluded when group membership was determined in Anderssen et al. (2018).

In Section 2.2, we discussed the issue of CLI and possible overlap between the two languages. As shown, there is generally no overlap between Norwegian and English with respect to SS, except in one context with complete overlap. This is in (*yes/no-* and *wh-*) questions, where both languages display verb movement: Norwegian has V2 word order and English has subject-auxiliary inversion. Although both Norwegian and English allow both word orders, the two languages

display different preferences, Norwegian for shifted (S-Neg) and English for unshifted word order (Neg-S). Out of the 145 main clause contexts for SS in the CANS corpus, there are 54 questions, 29 with S-Neg and 25 with Neg-S word order. This means that questions make up a considerably larger proportion of the unshifted cases, 44.8% (26/58) vs. only 33.3% (29/87) of the shifted cases. This also means that almost half of all questions (46.3%, 25/54), appear with Neg-S, while only about a third (36.3%, 33/91) of all declaratives appear with this unshifted word order; cf. the total numbers in Table 6. This indicates that CLI from English may be playing a role here, in that there is less SS in contexts where there is structural overlap with English (i.e. in questions). An example of the unshifted word order in a tag question is provided in (25). Despite the fact that unshifted word order is occasionally possible in non-HN, the word order in this example does seem odd and is presumably due to influence from English.

(25) *ja* ^{##} *er rart   hvor fort disse åra   har gått forbi er **ikke det**?*
      yes   is strange how fast these years have gone by   is not it
      "Yes, it is strange how fast these years have gone by, isn't it?"

In Section 3, we predicted that there should be a difference between the two groups of HSs with respect to influence from English: More specifically, we expected proficient speakers (i.e. the speakers in the *Norwegian* group) to be affected by CLO and thus overuse SS (the structure that is not preferred in English), while the speakers in the *English* group should be affected by CLI and overuse non-shifted order (the structure that *is* preferred in English). Table 6 provides an overview of declaratives and questions with shifted and unshifted word order produced by the two groups of speakers (*Norwegian* group, $n = 21$, and *English* group, $n = 7$), plus the 22 remaining speakers simply for completeness. Unfortunately, the raw numbers are now quite low, with the *Norwegian* group producing 75 relevant Examples (26 questions and 49 declaratives) and the *English* group only 18 (9 questions and 9 declaratives). Nevertheless, it seems clear that our predictions are *not* borne out: While the *Norwegian* group produces SS in declaratives at a level not too far from non-heritage Norwegians (71.4%), they produce considerably less SS in questions, where there is overlap with English, only 53.5%. This result is the opposite of what we expected for this group, indicating that the speakers are affected by CLI and not CLO. With respect to the *English* group, the raw numbers are so low that it is impossible to draw any firm conclusions: The total numbers show that there is in fact more SS in questions than in declaratives, a very surprising result. However, a closer look at the data shows that the 18 examples are produced by 6 different speakers, with only a couple of examples each, meaning that there is virtually no variety of contexts produced by individual speakers (one speaker produces only questions, another only declaratives, etc.).

Table 6. Overview of shifted (S-Neg) vs. unshifted (Neg-S) word order in questions and declaratives, divided into speaker groups (CANS, *n* = 50)

Speaker groups	Questions		Declaratives	
	S-Neg	Neg-S	S-Neg	Neg-S
Norwegian (*n* = 21)	**53.8% (14/26)**	46.2% (12/26)	**71.4% (35/49)**	28.6% (14/49)
English (*n* = 7)	66.6% (6/9)	33.3% (3/9)	55.6% (5/9)	44.4% (4/9)
Remaining (*n* = 22)	47.4% (9/19)	52.6% (10/19)	54.5% (18/33)	45.5% (15/33)
Total	**53.7% (29/54)**	**46.3% (25/54)**	**63.7% (58/91)**	**36.3% (33/91)**

We thus seem to have a situation where *all* speakers are somewhat affected by the similarity with an overlapping structure in English. In order to investigate that further, we have checked the verb types used with shifted and unshifted word orders, since subject-auxiliary inversion in English only appears with *be* and auxiliaries, not with lexical verbs. The copula and auxiliaries are quite common in the corpus, making up as much as 66.9% (97/145) of all verbs in main clauses, more specifically 81.5% (44/54) of the questions and 58.2% (53/91) of the declaratives. Although raw numbers again become very low when we break down the data into verb types, Table 7 shows that lack of SS (Neg-S word order) is somewhat more frequent with auxiliaries and *be* than with lexical verbs, i.e. in situations where negation in English can be attached to the verb as a clitic.

Table 7. Overview of shifted (S-Neg) vs. unshifted (Neg-S) word order in questions and declaratives, divided into verb types (CANS, *n* = 50)

Verb types	Questions		Declaratives	
	S-Neg	Neg-S	S-Neg	Neg-S
Lexical verbs	70% (7/10)	30% (3/10)	73.7% (28/38)	26.3% (10/38)
Be/aux	50% (22/44)	50% (22/44)	56.6% (30/53)	45.4% (23/53)
Total	**53.7% (29/54)**	**46.3% (25/54)**	**63.7% (58/91)**	**36.3% (33/91)**

In questions with lexical verbs (10/54), SS appears 70% (7/10), while in questions with *be* or auxiliaries, the rate of SS is only 50% (22/44). Interestingly, we find the same pattern in declaratives, with SS appearing 73.7% (28/38) with lexical verbs, but only 56.6% (30/53) with *be* and auxiliaries. These results thus further indicate that Neg-S word order may be affected by English, more specifically the frequent contraction of *be*/aux and negation in English (*isn't, hasn't*, etc.). With the caveat that numbers are quite low, we may also conclude that there does not seem to be any difference between the speaker groups with respect to the influence from English in the case of SS.

5.2 Object shift in heritage Norwegian

An examination of OS in the CANS corpus first and foremost reveals that the heritage speakers shift pronominal objects with nominal and non-nominal antecedents at 61% and 11% (Table 8). These results suggest that the heritage speakers distinguish between objects with nominal and non-nominal antecedents and tend to shift the former and not the latter, as illustrated in (26)–(27). Nevertheless, the distribution of the shifted and non-shifted position is significantly different from that of non-heritage speakers: For heritage speakers, 61% of pronominal objects with nominal antecedents shift, while for Norwegian speakers in Norway the equivalent proportion is 87% ($p < 0.001$).

Table 8. Pronominal objects with nominal and non-nominal antecedents (CANS, $n = 50$)

	O-Neg	Neg-O
Nominal antecedents	61% (25/41)	39% (16/41)
Non-nominal antecedents	11% (11/100)	89% (89/100)

(26) *Nå prøver jeg å finne et ord, men jeg finner **det ikke.***
now try I to find a word but I find it not.
"I'm trying to find a word, but I can't."

(27) a. *Snakker dere norsk med hverandre, eller?*
speak you Norwegian with each other or
"Do you speak Norwegian together?"

    b. *Nei, vi bruker **ikke det** noe mye, nei.* (that=speak Norwegian)
no we use not that any much no
"No, not very often."

With regard to the two groups of speakers, the numbers are too small to get a clear picture of possible differences between them. Indeed, there are many speakers in both groups who do not produce any relevant contexts. The *Norwegian* group has the lowest proportion of speakers of this type (14.3%, 3/21), while the *English* group has a somewhat higher proportion (28.6%, 2/7). Interestingly, the group consisting of the 22 speakers who did not produce enough relevant structures to be included in the analysis in Anderssen et al. (2018) has the most speakers who produce no relevant contexts (45.5%, 10/22), which is in line with the results from the previous study.

    Table 9 provides an overview of the production of OS with pronouns with nominal and non-nominal antecedents in the two groups; the remaining 22 speakers have also been included to make the picture complete. The two groups are very similar when it comes to the proportion of object pronouns with nominal antecedents that shift. However, note that there are very few examples in the *English* group, and

the seven examples are produced by four of the seven speakers. With respect to pronominal objects with non-nominal antecedents, there is a clear difference between the two groups, with the *Norwegian* group shifting these pronominal objects to a similar extent to (6.5%) and the *English* group shifting them somewhat more (27%) than in homeland Norwegian (5%). It is difficult to know how to interpret these results. However, the total results seem to suggest that there is a general erosion taking place, whereby OS is gradually being lost. Interestingly, when it comes to objects with non-nominal antecedents, the speakers in the more proficient group behave in a manner very similar to speakers in Norway. For the less proficient group the tendency is in the opposite direction; they shift pronominal objects with non-nominal antecedents at a higher rate than homeland Norwegian speakers. This development seems to be one towards increasing indeterminacy in the grammar with both types of object pronouns.

Table 9. Overview of shifted (O-Neg) vs. unshifted (Neg-O) word order with pronominal objects with nominal and non-nominal antecedents, divided into speaker groups (CANS, $n = 50$)

Speaker groups	Nominal antecedents		Non-nominal antecedents	
	O-Neg	Neg-O	O-Neg	Neg-O
Norwegian ($n = 21$)	57.7% (15/26)	42.3% (11/26)	6.5% (3/46)	93.5% (43/46)
English ($n = 7$)	57.1% (4/7)	42.9% (3/7)	27% (4/15)	73% (11/15)
Remaining ($n = 22$)	75% (6/8)	25% (2/8)	10.3% (4/39)	89.7% (35/39)
Total	**61% (25/41)**	**39% (16/41)**	**11% (11/100)**	**89% (89/100)**

6.   Discussion

In Section 3, we formulated a number of predictions, repeated here for convenience:

(21′) If frequency plays a role (in local domains):
  a.   OS should be vulnerable (non-shifted position more frequent than shifted position)
  b.   SS should be unaffected in main and embedded clauses (shifted position more frequent than unshifted)

(22′) If cross-linguistic similarities/differences play a role in contexts with total overlap:
  a.   Proficient speakers should prefer SS (affected by CLO)
  b.   Less proficient (attrited) speakers should prefer non-shifted order (CLI)

From the results reported in the previous section, it is clear that there is no global frequency effect in the HN data. The fact that OS is massively less frequent than SS and that embedded clauses are considerably less frequent than main clauses does not seem to play any role. OS is not more affected than SS; in fact, both constructions seem to be equally vulnerable. Embedded clauses are not more affected than main clauses; the results actually show the opposite: While SS in main clauses is attested significantly less than in non-HN, embedded clauses are unproblematic. We believe that this may be due to a general preference for subject-initial clauses in both languages.

If we consider frequency in more local domains (i.e. each construction or clause type separately), the predictions in (21) also do not seem to be straightforwardly borne out, since both SS and OS seem to be affected to the same extent. The fact that pronominal objects in HN shift across negation significantly less than in non-HN (61% vs. 87%) could in principle be argued to be due to the high frequency of unshifted compared to shifted objects. That is, the HN speakers may have developed a preference for the more frequent word order for objects, the unshifted one (Neg-O). Nevertheless, the rate of shifted objects is relatively high. Furthermore, a preference based on local frequency should lead to the opposite result for SS: Since most subjects are pronouns and most pronouns shift, we would expect SS to be unaffected, which it is not (60.0% vs. 84.7% and 87.9% in non-HN). Thus, an explanation in terms of frequency alone does not hold.

Anderssen et al. (2018) are unable to determine whether the higher use of postnominal possessives and the suffixal article in modified definites is due to frequency or cross-linguistic overcorrection (CLO). This is also true of the noun-adjective order discussed in Kupisch's (2014) original study. Thus, for all the three phenomena that have been explained with reference to CLO, the structure that has been overcorrected has also been the more frequent option in the heritage language. Our study might shed some light on this issue, because there is no superficial structural similarity in the case of OS and SS in general, and CLO should not be available, as there simply is no structure to inhibit. The results reported here are consequently important for two reasons, (i) they reveal that frequency alone has a limited effect in heritage language, and (ii) they provide (indirect) support for CLO as a factor in heritage languages. Note, however, that this does not preclude the possibility that the two factors have to work in tandem, and that CLO only affects structures that are also more frequent.

A complicating factor when it comes to the effect of frequency on SS in HN is related to the final set of predictions in (22). According to these predictions, the more proficient speakers should be affected by CLO and overuse SS in questions, while the less proficient speakers should be affected by CLI and prefer the more

English-like word order. This prediction does not seem to have been borne out, as there is considerably less SS in questions in HN, and this difference is especially pronounced in the *Norwegian* group, where there is a higher rate of SS in declaratives than in questions (71.4% vs. 53.8%). It thus appears that in situations with complete structural overlap, that is, structures where both languages display the same type of word order variation, CLO does not apply. Instead, both more and less proficient speakers are affected by CLI and influenced by the word order that is preferred in the dominant language. Furthermore, there is less SS also in other contexts, especially with auxiliaries and *be*, where the negation would tend to be cliticized onto the verb in English (e.g. *isn't, hasn't*), resulting in (aux/be)-Neg-S word order. This indicates that the HSs are influenced by the more common order in their dominant language. However, returning to the question of (local) frequency, it appears that even though the rate of SS is to some extent affected by CLI in questions (causing subjects not to shift), especially in the *Norwegian* group, this cannot entirely explain the low rate of SS. Recall that we predicted that SS should be robust because subjects usually shift, while OS should be vulnerable because most objects do not undergo OS, but what we found was that subjects and objects shift at similar rates (60.0% vs. 61.0%). It thus seems that the effect of CLI is not strong enough to explain why frequency has such a limited effect on SS, because even if we only consider declaratives, where there should be no CLI, subjects still only shift at 63.7%. Thus, the overall difference between SS and OS is surprisingly small, suggesting that frequency has a limited effect on the linguistic production of these HSs, even when the effect of CLI is taken into account.

Several recent studies have shown that both languages of a bilingual always stay active (e.g. Hartsuiker, Pickering & Veltkamp 2004; Martin, Dering, Thomas & Thierry 2009) and that speakers need to inhibit the other language in monolingual situations. This should be more difficult when the dominant language is the one that has to be inhibited (Sorace 2011), which is the case when these speakers use Norwegian, their heritage language. Against this backdrop, we may postulate that there is a difference between partial and total overlap and more and less proficient speakers: In cases of partial overlap, when a choice has to be made between two word orders in HN, the only possible word order in English is simultaneously activated and will need to be inhibited. In this case, a lower proficiency in Norwegian will cause a speaker to be less successful at inhibiting the dominant language and thus be influenced by the overlapping word order (CLI), cf. Table 3 in Section 2.4. In possessives, this means that POSS-N gets overused (by speakers with a lower proficiency) and in modified definites that the prenominal determiner is used. A speaker with a higher proficiency will successfully inhibit the English word order, in fact to the extent that the similar Norwegian word order is also inhibited, and because of this, the speaker will be more sensitive to the non-overlapping

(*different*) alternative (N-POSS) and overuse this word order (CLO) (cf. Table 4). Similarly, in modified definites, it is the suffixal article that is preferred, i.e. the structure that is more typically Norwegian. As mentioned above, it is possible that this effect will be exacerbated if the relevant structure is also more frequent than the alternative, which is the case with both N-POSS and the suffixal article. Thus, if inhibition of the dominant language is weaker, the result is CLI, while if it is stronger, the result is CLO.

In cases of total overlap, on the other hand, both word orders in English will be activated – together with the corresponding preferences. This means that the two options in HN are in direct competition with corresponding structures in English, and one of them will be preferred and more frequent in the dominant language and thus harder to inhibit (i.e. the one that is dispreferred and less frequent in Norwegian). In such a situation, both word orders require inhibiting. This has two consequences: (i) no CLO can take place, because there is no typical Norwegian structure that is different from English that could be overcorrected to, and (ii) what has to be inhibited is not only the two word orders, but also the preference for one of them. This seems to be harder and require more resources than just inhibiting one overlapping structure. This could explain why we find CLI in both groups of speakers of HN in SS structures. In fact, the effect of CLI is particularly clear in the *Norwegian* group, probably because they are the only ones proficient enough to produce a critical number of contexts for SS (non-subject-initial clauses).[5] Table 10 illustrates the situation with complete overlap schematically; the two overlapping structures are shaded in English and the option that is affected by CLI is given in italics (again, language dominance is indicated by boldface):

Table 10. CLI from English into HN in structures with complete overlap (both groups)

Language\Structure	Questions		Questions	
Norwegian	Q V	S NEG (preferred)	*Q V*	*NEG S*
English	**Q AUX/BE**	**S NEG**	**Q AUX/BE**	**NEG S (preferred)**

To summarize, we have seen that, contrary to expectations, there are limited frequency effects in HN, as OS and SS seem to be equally vulnerable to attrition. However, as CLI is an additional factor at play in some contexts requiring SS (questions), the comparison between SS and OS is not straightforward. In cases of total structural overlap where the two languages have opposite word order preferences,

---

5.   Note that we are talking about speakers where even the most proficient ones have very few opportunities to use the heritage language. Speakers who use both languages on a regular basis are generally expected to successfully inhibit the dominant language without any of these effects.

we have shown that all speakers seem to be affected by CLI. However, the difference between the contexts with superficial structural overlap (questions) and those without such overlap (declaratives) is more pronounced in the more proficient *Norwegian* group.

7.   Summary and conclusion

This paper set out to investigate how the factors frequency and structural similarity/difference play out in Heritage Norwegian for two word order phenomena, Subject Shift (SS) and Object Shift (OS), i.e. contexts where certain subjects and objects may move across negation. In both cases, there are major differences in the frequencies of the two word orders; subjects typically precede and objects typically follow negation. Previous research on structures with partial overlap between Norwegian and English has indicated that highly proficient heritage speakers are affected by frequency and cross-linguistic overcorrection (CLO, Kupisch 2014), while less proficient speakers are sensitive to cross-linguistic similarities between the two languages (CLI). Given that there is no structural overlap between English and Norwegian in these constructions, we do not expect CLI or CLO, except in questions with auxiliaries or *be*, where there is complete overlap between English and Norwegian – with opposite preferences. Our findings show that OS is somewhat vulnerable, and in isolation, this could possibly be due to low frequency. However, SS is also vulnerable in main clauses, even though this word order is clearly more frequent. This is partly explained by complete overlap between English and Norwegian. Contrary to expectations, the more proficient speakers are not affected by CLO and do not overuse SS in these structures; rather, both proficient and less proficient speakers overuse the unshifted word order, i.e. the one preferred in English. We thus argue that, in situations with complete structural overlap, where both word orders have to be inhibited in the dominant language, all heritage speakers may be affected by CLI.

Acknowledgements

Funding information

References

Anderssen, M. & Bentzen, K. 2012. Norwegian object shift as IP-internal topicalization. *Nordlyd* 39.1: *The Grammar of Objects*, 1–23.

Anderssen, M. & Westergaard, M. 2010. Frequency and economy in the acquisition of variable word order. *Lingua* 120: 2569–2588. https://doi.org/10.1016/j.lingua.2010.06.006

Anderssen, M., Lundquist, B. & Westergaard, M. 2018. Cross-linguistic similarities and differences in bilingual acquisition and attrition: Possessives and double definiteness in Norwegian heritage language. *Bilingualism: Language and Cognition* 21(4): 748–764. https://doi.org/10.1017/S1366728918000330

Anderssen, M., Bentzen, K. & Rodina, Y. 2012. Topicality and complexity in the acquisition of Norwegian object shift. *Language Acquisition* 19(1): 39–72. https://doi.org/10.1080/10489223.2012.633844

Anderssen, M., Bentzen, K., Rodina, Y. & Westergaard, M. 2010. The acquisition of apparent optionality: The word order of subject and object shift constructions in Norwegian. In *Variation in the Input: Studies in the Acquisition of Word Order*, M. Anderssen, K. Bentzen & M. Westergaard (eds), 241–270. Dordrecht: Springer. https://doi.org/10.1007/978-90-481-9207-6_10

Anderssen, M., Rodina, Y., Mykhaylyk, R. & Fikkert, P. 2014. The acquisition of the dative alternation in Norwegian. *Language Acquisition* 21: 72–102. https://doi.org/10.1080/10489223.2013.864296

Anderssen, M. 2006. The Acquisition of Compositional Definiteness in Norwegian. PhD dissertation, University of Tromsø.

Andréasson, M. 2008. Not all objects are born alike – accessibility as a key to pronominal object shift in Swedish and Danish. In *Proceedings of the LFG08 Conference*, M. Butt & T. Halloway King (eds), 26–45. Stanford CA: CSLI.

Andréasson, M. 2010. Object shift or object placement in general. In *Proceedings of the LFG10 Conference*, M. Butt & T. Halloway King (eds), 26–42. Stanford CA: CSLI.

Bentzen, K. 2009. Subject positions and their interaction with verb movement. *Studia Linguistica* 63(3): 1–31. https://doi.org/10.1111/j.1467-9582.2009.01162.x

Bentzen, K. 2014. Object shift. *Nordic Atlas of Language Structures (NALS)* 1: 332–343.

Bentzen, K., Anderssen, M. & Waldmann, Ch. 2013. Object shift in Mainland Scandinavian: A corpus study of Danish, Norwegian and Swedish. *Nordic Journal of Linguistics* 36(2): 115–151. https://doi.org/10.1017/S0332586513000218

Cardinaletti, A. 2004. Towards a cartography of subject positions. In *The Structure of CP and IP: The Cartography of Syntactic Structures*, Vol. 2, L. Rizzi (ed.), 115–165. Oxford: OUP.

Erteschik-Shir, N. 2005. Sound patterns of syntax: Object shift. *Theoretical Linguistics* 31(1–2): 47–94.

Hartsuiker, R. J., Pickering, M. J. & Veltkamp, E. 2004. Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science* 15(6): 409–414. https://doi.org/10.1111/j.0956-7976.2004.00693.x

Haugen, E. 1953. *The Norwegian Language in America. A Study in Bilingual Behavior*. Bloomington IN: Indiana University Press. https://doi.org/10.9783/9781512820522

Holmberg, A. 1986. Word Order and Syntactic Features in the Scandinavian Languages and English. PhD dissertation, University of Stockholm.

Holmberg, A. 1999. Remarks on Holmberg's generalization. *Studia Linguistica* 53(1): 1–39. https://doi.org/10.1111/1467-9582.00038

Jayaseelan, K. 2001. IP-internal topic and focus phrases. *Studia Linguistica* 55(1): 39–75. https://doi.org/10.1111/1467-9582.00074

Johannessen, J. B. 2015. The Corpus of American Norwegian Speech (CANS). In *Proceedings of the 20th Nordic Conference of Computational Linguistics, NODALIDA 2015*, B. Megyesi (ed.), 297–300. Linköping: University Electronic Press.

Johannessen, J. B. & Laake, S. 2017. Norwegian in the American Midwest: A common dialect? *Journal of Language Contact* 10: 5–21. https://doi.org/10.1163/19552629-01001002

Johannessen, J. B. & Lake, S. 2012. Østnorsk som fellesdialect i Midtvesten. *Norsk Lingvistisk Tidsskrift* 30(2): 365–380.

Johannessen, J. B. & Salmons, J. 2015. The study of Germanic heritage languages in the Americas. In *Germanic Heritage Languages in North America* [Studies in Language Variation 18], J. B. Johannessen & J. Salmons (eds), 1–17. Amsterdam: John Benjamins. https://doi.org/10.1075/silv.18.001int

Josefsson, G. 2010. Object shift and optionality: An intricate interplay between syntax, prosody and information structure. *Working Papers in Scandinavian Syntax* 86: 1–24.

van Kemenade, A. & Los, B. 2006. Discourse adverbs and clausal syntax in Old and Middle English. In *The Handbook of the History of English*, A. van Kemenade & B. Los (eds), 224–248. Malden MA: Blackwell. https://doi.org/10.1002/9780470757048.ch10

Köpke, B. & Schmid, M. 2004. Language attrition: The next phase. In *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues* [Studies in Bilingualism 28], M. Schmid, B. Köpke, M. Keijzer & L. Weilemar (eds), 1–43. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.28.02kop

Kupisch, T. 2013. A new term for a better distinction? A view from the higher end of the proficiency scale. *Theoretical Linguistics* 39(3–4): 203–214.

Kupisch, T. 2014. Adjective placement in simultaneous bilinguals (German-Italian) and the concept of cross-linguistic overcorrection. *Bilingualism: Language and Cognition* 17(1): 222–233. https://doi.org/10.1017/S1366728913000382

Kupisch, T. & Rothman, J. 2018. Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism* 22(5): 564–582.

Lødrup, H. 2011. Norwegian possessive pronouns: Phrases, words or suffixes? In *Proceedings of the LFG11 Conference*, M. Butt & T. Holloway King (eds), 383–403. Stanford CA: CSLI.

Lohndal, T. & Westergaard, M. 2016. Grammatical gender in American Norwegian heritage language: Stability or attrition? *Frontiers in Psychology* 7.344. https://doi.org/10.3389/fpsyg.2016.00344

Lovoll, O. S. 1999. *The Promise of America: A History of the Norwegian-American People*, revised edn. Minneapolis MN: University of Minnesota Press.

Martin, C. D., Dering, B., Thomas, E. M. & Thierry, G. 2009. Brain potentials reveal semantic priming in both the 'active' and the 'non-attended' language of early bilinguals. *NeuroImage* 47(1): 326–333. https://doi.org/10.1016/j.neuroimage.2009.04.025

Mikkelsen, L. 2011. On prosody and focus in object shift. *Syntax* 14(3): 230–264. https://doi.org/10.1111/j.1467-9612.2011.00152.x

Mohr, S. 2005. *Clausal Architecture and Subject Positions: Impersonal Constructions in the Germanic Languages* [Linguistik Aktuell/Linguistics Today 88]. Amsterdam: John Benjamins. https://doi.org/10.1075/la.88

Montrul, S. 2002. Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition* 5: 39–68. https://doi.org/10.1017/S1366728902000135

Montrul, S. 2008. *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor* [Studies in Bilingualism 39]. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.39

Montrul, S. 2014. Incomplete L1 acquisition. In *The Cambridge Handbook of Second Language Acquisition*, J. Herschensohn & M. Young-Scholten (eds), 353–371. Cambridge: CUP.

Müller, N. & Hulk, A. 2001. Cross-linguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition* 4: 1–21. https://doi.org/10.1017/S1366728901000116

Nilsen, Ø. 1997. Adverbs and A-shift. *Working Papers in Scandinavian Syntax* 59: 1–31.

NoTa – Norsk Talespråkskorpus, Oslodelen [Corpus of Spoken Norwegian, the Oslo part], Tekstlaboratoriet, ILN, University of Oslo.

Polinsky, M. 2006. Incomplete acquisition: American Russian. *Journal of Slavic Linguistics* 14: 191–262.

Putnam, M. & Sánchez, L. 2013. What's so incomplete about incomplete acquisition? A prolegomenon to modeling heritage grammars. *Linguistic Approaches to Bilingualism* 5(2): 478–508. https://doi.org/10.1075/lab.3.4.04put

Radford, A. 1992. The acquisition of the morphosyntax of finite verbs in English. In *The Acquisition of Verb Placement: Functional Categories and V2 Phenomena in Language Acquisition*, J. Meisel (ed.), 23–62. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-011-2803-2_2

Roeper, T. 1999. Universal bilingualism. *Bilingualism: Language and Cognition* 2(3): 169–186. https://doi.org/10.1017/S1366728999000310

Roeper, T. 2007. What frequency can do and what it can't. In *Frequency Effects in Language Acquisition: Defining the Limits of Frequency as an Explanatory Concept*, I. Gülzow & N. Gagarina (eds), 23–48. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110977905.23

Scontras, G., Fuchs, Z. & Polinsky, M. 2015. Heritage language and linguistic theory. *Frontiers in Psychology* 6. 1545. https://doi.org/10.3389/fpsyg.2015.01545

Simonsen, H. G. 1990. Barns fonologi: system og variasjon hos tre norske og et samoisk barn. PhD dissertation, University of Oslo.

Snyder, W. 2007. *Child Language: The Parametric Approach*. Oxford: OUP.

Sorace, A. 2011. Pinning down the concept of "interface" in bilinguals. *Linguistic Approaches to Bilingualism* 1(1): 1–33. https://doi.org/10.1075/lab.1.1.01sor

Svenonius, P. 2002. Subject positions and the placement of adverbials. In *Subjects, Expletives and the EPP*, P. Svenonius (ed.), 201–242. Oxford: OUP.

Venås, K. 1971. Om posisjonen til neksusadverbialet *ikkje*. *Maal og Minne* 3–4: 124–173.

Vikner, S. 2006. Object shift. In *The Blackwell Companion to Syntax*, Vol 3, M. Everaert & H. van Riemsdijk (eds), 392–436. Oxford: Blackwell. https://doi.org/10.1002/9780470996591.ch46

Vogel, R. 2006. Weak function word shift. *Linguistics* 44(5): 1059–1097.
https://doi.org/10.1515/LING.2006.034

Westergaard, M. 2008. Verb movement and subject placement in the acquisition of word order: pragmatics or structural economy? In *First Language Acquisition of Morphology and Syntax: Perspectives across Languages and Learners* [Language Acquisition and Language Disorders, 45] P. Guijarro-Fuentes, P. Larranaga & J. Clibbens (eds), 61–86. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.45.03wes

Westergaard, M. 2009a. Microvariation as diachrony: A view from acquisition. *Journal of Comparative Germanic Linguistics* 12: 49–79. https://doi.org/10.1007/s10828-009-9025-9

Westergaard, M. 2009b. *The Acquisition of Word Order: Micro-cues, Information Structure and Economy* [Linguistik Aktuell/Linguistics Today 145]. Amsterdam: John Benjamins.
https://doi.org/10.1075/la.145

Westergaard, M. 2011. Subject positions and information structure: The effect of frequency on acquisition and change. *Studia Linguistica* 3: 299–332.
https://doi.org/10.1111/j.1467-9582.2011.01183.x

Westergaard, M. 2014. Linguistic variation and micro-cues in first language acquisition. *Linguistic Variation* 14(1): 26–45. https://doi.org/10.1075/lv.14.1.02wes

Westergaard, M. & Anderssen, M. 2015. Word order variation in Norwegian possessive constructions: bilingual acquisition and attrition. In *Germanic Heritage Languages in North America: Acquisition, Attrition and Change* [Studies in Language Variation 18], J. B. Johannessen & J. Salmons (eds), 21–45. Amsterdam: John Benjamins. https://doi.org/10.1075/silv.18.01wes

Westergaard, M. & Bentzen, K. 2007. The (non-) effect of input frequency on the acquisition of word order in Norwegian embedded clauses. In *Frequency Effects in Language Acquisition: Defining the Limits of Frequency as an Explanatory Concept*, I. Gülzow & N. Gagarina (eds), 271–306. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110977905.271

Westergaard, M. & Lohndal, T. 2019. Verb second word order in Norwegian heritage language: Syntax and pragmatics. In *Variable Properties in Language: Their Nature and Acquisition*, D. Lightfoot & J. Havenhill (eds), 91–102. Washington DC: Georgetown University Press.

Westergaard, M., Vangsnes, Ø.A. & Lohndal, T. 2017. Variation and change in Norwegian wh-questions: The role of the complementizer som. *Linguistic Variation* 17(1): 8–43.
https://doi.org/10.1075/lv.17.1.02wes

Language contact

Gender agreement in Spanish L2 learners and heritage speakers

Jessica Diebowski

Bergische Universität Wuppertal

This paper examines the effect of language contact on the knowledge of Spanish gender assignment and agreement in adult second language learners and simultaneous bilinguals (heritage speakers of Spanish), all residing in the Geneva area of New York State. The data comes from 27 English-speaking learners of Spanish and 27 bilingual speakers, who completed a grammatical judgment task (GJT) and an oral elicitation production task (OPT). In particular, the paper investigates whether the successful acquisition of gender is dependent on the extent of exposure to the target language. The results show successful acquisition of gender assignment and agreement in all groups. In addition, the findings indicate that the extent to which Spanish is used seems to affect the gender accuracy of Spanish L2 learners as opposed to heritage speakers, who perform at ceiling in the GJT and oral task regardless of the frequency of heritage language (HL) use.

Keywords: gender assignment, gender agreement, L2 learners, heritage speakers, Spanish, amount of exposure

1. Introduction

Languages are dynamic entities that may change over time (Montrul 2006a, 2006b: 379; Thomason & Kaufman 1991; Ramírez 2003). In today's multilingual societies in which various types of language contact situations occur, such as second language acquisition and bilingual first language acquisition (Montrul 2006b: 379), changes at all levels of the linguistic system are noticeable (Thomason & Kaufman 1991). These changes may manifest in various forms, e.g. simplifications, avoidance of certain target language forms, preference for other non-target forms and transferring patterns from one language to another (cf. Kolehmainen, Meriläinen & Riionheimo 2014). However, the linguistic outcomes in languages involved in the contact situation may differ depending on numerous factors, among them the

types of speakers (i.e. here: Spanish heritage speakers and English-speaking L2 learners of Spanish), the heterogeneity among speakers, the availability of input and the intensity of contact with the language (cf. Montrul 2006a). This chapter discusses these outcomes through a comparison of gender accuracy in Spanish second language acquisition and Spanish heritage language acquisition. Its aim is to cast more light on the similarities and differences between Spanish L2 learners (foreign language learners) and Spanish heritage speakers (simultaneous bilinguals) at different levels of language contact in the domain of gender agreement. In the existing literature, it is often noted that the category of gender is one problem area common to both Spanish L2 learners and Spanish heritage speakers which typically manifests in comprehension and production, e.g. in non-target like realizations of gender assignment and agreement due to influence from English (Lipski 1993; Montrul 2002; Polinsky 2007; Silva-Corvalán 2003). In the case of L2 learners, who start learning the second language usually in a classroom setting after puberty, many researchers (e.g. Montrul, Foote & Perpiñán 2008) claim that the acquisition of gender is dependent upon the limits of ultimate attainment in the L2 gram-mar. These authors argue that L2 learners cannot attain native-like results due to maturational constraints. They also insist that L2 learners reveal a high number of errors in gender assignment and agreement, despite many years of studying the foreign language (e.g. Hawkins & Franceschina 2003, to mention only one example). Regarding Spanish heritage speakers living in the United States, it has been widely attested that gender assignment and agreement particularly in the DP is often affected (Alarcón 2008; Montrul 2002, 2004a, 2005; Silva-Corvalán 1994, 2003; Valdés 1995). Scholars like Lipski (1993, 1996), Montrul (2002), Montrul, Foote, & Perpiñán (2008), Silva-Corvalán (1994) and Toribio (2001) maintain that heritage speakers have undergone a loss of their language skills in their heritage language due to exposure to and use of the majority language (English). Montrul (2006b: 381) even goes one step further by considering the divergent outcome by heritage speakers from a postulated "monolingual baseline" (fusing the norm and linguistic behavior of monolingual speakers) as a case of

> incomplete acquisition because in many respects the grammars of the family lan-guage (in this case Spanish) look like the grammars of intermediate and advanced L2 learners of Spanish […] and not like grammars of monolingual Spanish speakers.

More recently, Pascual y Cabo & Rothman (2012) and Rothman & Treffers-Daller (2014) criticized the use of monolingual speakers as a benchmark against which bilinguals are measured. For bilingual speakers, two languages interact at many levels and require a great amount of processing costs, thus variants may arise from language contact at the individual and societal levels, which are rather different from monolinguals.

This chapter explores the effect of different degrees of language contact in second language acquisition and simultaneous bilingualism at an individual level on gender accuracy. More specifically, I pose the question of what happens to the knowledge of gender agreement when Spanish L2 learners and heritage speakers have regular, occasional, or seldom contact with the target language, i.e. Spanish spoken primarily in Latin and South America. This study will show that while Spanish L2 learners and Spanish heritage speakers are exposed to language contact at different levels, their knowledge of gender is not incomplete.

2. Theoretical considerations on language contact

Much of the existing literature on language contact and multilingualism is devoted to various fields of research such as contact linguistics, second language and bilingual language acquisition focusing on the manifold linguistic phenomena and taking different perspectives, e.g. language contact at the individual or societal level (cf. Filppula, Klemola & Paulasto 2009). Language contact at the individual level regards the person who is bi- or even multilingual as locus of language contact (cf. Weinreich 1953). However, individual bilingualism is both understood and defined differently by scholars. Bloomfield (1935) considers individuals to be bilingual when they are able to use two or more languages in a native-like way. Weinreich (1953: 5) opposes this view by contending that bilingualism is based on the alternative use of two languages, referring to the general dimension of a language. Language contact at the societal level regards the society as locus of language contact. This type of language contact can be identified within communities of speakers using two or more languages and interacting with each other.

What consequences does language contact have for speakers? Appel & Muysken (2005: 9) point out:

> Many people find themselves at the frontier of two languages. […] [T]here are many ways of coping with this situation. The structural characteristics of languages involved impose an outer limit on the possible outcomes of language contact.

In terms of the United States, where the majority language English comes into contact with Spanish as the minority language, little has been said about the effects of language contact on areas such as morphosyntax and inflectional morphology, which have been identified as being highly susceptible to attrition or loss in bilingual environments (Bolonyai 2007; Montrul 2002; Silva-Corvalán & Treffers-Daller 2016; Toribio 2001). This paper will deal with the dimension of language contact at an individual rather than a societal level. From now on, the term language contact at an individual level is meant to refer to the amount of language use of an individual.

3.   Gender in Spanish and English

This study adopts the definition of grammatical gender as agreement between the noun and other targets like adjectives and determiners (Corbett 1991; Hockett 1958). Although there is a recurrent claim in the generative literature that gender is a functional head in syntax, comparable to number (cf. Picallo 1991), another view is taken here. The assumption is made that grammatical gender is a lexical feature of nouns and thus part of the lexical entry of nouns (cf. Roca 1989 for Spanish; Dewaele & Véronique 2001 for French; Cantone & Müller 2008 for Italian; Eichler, Jansen & Müller 2013 for French, Italian and German). In the following, grammatical gender in Spanish and English is presented with respect to (in)definite articles and adjectives.

3.1   Spanish

Like other Romance languages, such as Italian and French, Spanish has a binary gender system which distinguishes between two grammatical gender classes: masculine and feminine. Animate nouns referring to people and large or domesticated animals typically have a masculine and a feminine form, as in (1). However, there are many animate nouns, especially animal names such as *orca* 'killer whale', *mosca* 'fly', *avispa* 'wasp' (feminine epicene) or *tiburón* 'shark', *papagayo* 'parrot' (masculine epicene), which could be either male or female. Thus, they need to be specified with either the adjective *macho* for male sex and *hembra* for female sex. In Spanish, inanimate nouns are arbitrarily classified as masculine or feminine, as in (2):

(1)   a.   gato
          cat.MASC
      b.   gata
          cat.FEM

(2)   a.   espejo
          mirror.MASC
      b.   camisa
          shirt.FEM

In Spanish, there are consistent morphological cues to the noun's gender. Masculine (Masc) nouns commonly end in the canonical inflectional morpheme /-o/ (*espejo* 'mirror'-MASC), whereas feminine (Fem) nouns end in the canonical inflectional morpheme /-a/ (*camisa* 'shirt'-FEM) (Green 1988). According to Teschner and Russell (1984), Spanish has a rather transparent gender system as 99.87% of the nouns ending in -o are masculine and 96.30% of the nouns ending in -a are feminine. Although the vast majority of nouns exhibit this so-called overt morphology

to indicate gender, the gender of some nouns cannot be predicted by following the above mentioned rules such as masculine *día* 'day'-Masc, or feminine *radio* 'radio'-FEM. Further exceptions are nouns ending in other vowels or consonants, which may be either masculine or feminine, e.g. /-e/ as in *valle* 'valley'-MASC or *fuente* 'fountain'-FEM, *césped* 'grass'-MASC or *sed* 'thirst'-FEM (cf. Sagarra & Herschensohn 2011).

Nouns and constituents such as definite/indefinite determiners, adjectives, complements and other elements that modify the noun, display gender marking (cf. Corbett 1991). Table 1 illustrates the definite and indefinite articles inflected for gender and number.

Table 1. Spanish gender paradigm for indefinite and definite articles (adapted from Bosque 2007)

Number	*Singular*		*Plural*	
Gender	*masc.*	*fem.*	*masc.*	*fem.*
Definite /	el	la	los	las
Indefinite article	un	una	unos	unas

Spanish adjectives also display gender marking, a procedure called agreement or concord (Carroll 1999; Zagona 2002). Most of the adjectives appear post-nominally and mirror the forms of the articles, i.e. they end in the final morpheme -o for masculine and -a for feminine, as shown in (3). However, some adjectives end in the morpheme -e and are not overtly inflected for gender and thus remain invariant, as *interesante* in (4):

(3) a. El   espejo blanco
           DET.MASC mirror white.MASC
           "The white mirror"
    b. La   camisa blanca
           DET.FEM shirt   white.FEM
           "The white shirt"

(4) a. el   libro interesante
           DET.MASC book interesting Ø
           "The interesting book"
    b. la   película interesante
           DET.FEM movie   interesting Ø
           "The interesting movie"

Let us now compare the Spanish gender system with the English one, which has a pronominal gender system.

3.2   English

The gender system in English differs significantly from Spanish, given that it lost grammatical gender in its transition from Old to Middle English (Cruzan 2003). Due to this development, the Modern English gender system became a pronominal gender system, in which semantic gender is only evident on pronouns and some nouns, such as *actor/actress* (Corbett 1991: 12). Gender concord between the noun and its accompanying elements such as articles, possessives, numerals, and adjectives no longer exists, since there are no inflectional markings. Hence, gender assignment is largely limited to nouns that refer to humans and animate beings, and is based on the sex of the referent, as in (5) and (6).

(5)   Masculine   Feminine
      a. *man*   b. *woman*

(6)   Masculine   Feminine
      a. *actor*   b. *actress*

In nouns referring to things, gender is assigned arbitrarily. Ships, boats and nations, for example, are considered to be feminine (Jarvis & Pavlenko 2008: 133). In English, as in Spanish, adjectives can be used both predicatively and attributively. In the former case, English and Spanish are similar in the placement of predicative adjectives, i.e. post-nominal following the copula (cf. Jaensch 2009: 123), as in (7). In the latter case, in English attributive adjectives are usually prenominal. However, the possibility of placing the adjective after the noun also exists, such as in *stars visible, members present* etc. In Spanish most attributive adjectives are placed post-nominally (8), while the possibility to place the adjective before the noun also exists and usually involves a change in the meaning of the adjective.

(7)   a.   The   chair   is white.
           The-Ø chair-Ø is white-Ø
      b.   La   silla   es blanca.
           The.FEM chair.FEM is white.FEM
           "The chair is white."

(8)   a.   the   white   chair
           the-Ø white-Ø chair-Ø
      b.   la   silla   blanca
           the.FEM chair.FEM white.FEM
           "the white chair"

In short, we have seen that gender assignment is lexical. In Spanish, gender is manifested syntactically through agreement in the noun phrase, as shown in (8b), and verb phrase, as in (7b). The Examples (7a) and (8a) show that English does not

have a noun classification system based on gender like Spanish (cf. Montrul, Foote & Perpiñán 2008: 511). Furthermore, English lacks overt gender marking on inanimate and animate nouns and other elements such as determiners and adjectives. Thus, English-speaking learners of Spanish find it challenging to acquire grammatical gender, a feature absent in their L1. However, the acquisition of gender has "a prestigious position in gender languages, and errors are very noticeable to native speakers as an indication of non-nativeness" (Herschensohn & Arteaga 2015: 216). The next section discusses studies investigating the acquisition of gender agreement in second and heritage language acquisition.

4.   Previous research

The acquisition of gender in various languages and language combinations (e.g. German, French, Spanish, Russian etc.) has been a topic of scholarly interest and investigation for some time. There is a large array of studies investigating gender in both monolingual and bilingual child language acquisition, in adult L2 and in heritage language acquisition. This section presents previous studies that address adult L2 learners and heritage speakers of Spanish.

Regarding the nominal morphology, Finnemann (1992) studied the speech of three first year college students over a period of six months. In an interview, the students were asked to establish gender agreement in the Spanish noun phrase with overt and non-overt noun morphology. She found that students display lower error rates with nouns showing the inflectional morphemes (-o and -a) than with the ones that are not overtly marked for gender. Finnemann attributed her findings to the effect of morphology on gender agreement accuracy, suggesting that overt noun endings help students to produce the correct gender. In a similar study, Fernández-García (1999) investigated the effect of morphology on gender agreement accuracy. Her results were comparable to the ones from Finnemann (1992). The participants produced higher accuracy rates with nouns that have overt gender markings than those that are not overtly marked. Furthermore, she reported that the participants of the study tend to change the non-overt noun ending into an overt ending, for instance in *la clarineta vs. el clarinet* 'the clarinet'-MASC. Previous empirical findings by Cain, Weber-Olsen & Smith (1987) also report that the participants alter the morphology of a non-overt noun ending into an overt one. Franceschina (2001) studied the speech of two Italian (aged 71 and 73) and two English speaking learners (aged 50 and 55), who were first exposed to Spanish in post-puberty. Her results showed similarities to the findings of Finnemann (1992) and Fernández-García (1999) where there was more accuracy with overt than non-overt gender marking on the nouns. Concerning the effect of proficiency,

Alarcón (2006) studied 139 English speakers of three different proficiency levels in Spanish. She tested gender assignment and agreement in a written grammar test. The subjects were asked to provide the correct definite article and adjective based on the given noun. The results indicated that proficiency level has a significant effect on gender accuracy. Advanced learners achieve a high accuracy rate on gender agreement. A study by Gess and Herschensohn (2001) examining the gender acquisition in Spanish by French speakers supports this finding. They reported that advanced French-speaking learners of Spanish achieved a high accuracy on gender agreement, while learners at a lower proficiency level are more inaccurate (cf. also White 2003: 137).

There are very few studies that investigate the differences in the number of errors and error patterns between intermediate and advanced L2 learners compared to heritage speakers. Martínez-Gibson (2011) conducted a comparative study on gender agreement errors in the spoken Spanish of heritage and L2 learners. She interviewed a total of 44 participants. They were asked to describe a picture which was presented to them. The results of the study revealed that the L2 learners are more inaccurate in gender assignment and agreement than the heritage speakers. Furthermore, she noted a higher rate of inaccuracy in noun-adjective gender agreement than in article-noun agreement. This finding is in accordance with previous studies carried out by Finnemann (1992) and Fernández-García (1999), who attribute this high number of errors in noun-adjective agreement to the fact that L2 learners acquire adjective agreement later than article-noun agreement. Earlier findings by Montrul, Foote and Perpiñán (2008) show similar results. Chini (1995) investigated the development of correct gender use in Italian L2 and found that articles preceded adjectives in correct gender agreement. These results are in accordance with findings on L1 acquisition (Tucker, Lambert & Rigault 1977) concerning the importance of determiners in the characterization of gender and suggest that in Spanish L1 and L2 acquisition the gender of nouns is learned primarily through determiners.

Having a closer look at the tools assessing L2 learners' and heritage speakers' knowledge of gender agreement, scholars found that Spanish L2 learners are more target-like than heritage speakers in written comprehension tasks. However, heritage speakers are more target-like in oral production tasks (see Montrul 2004b, 2005; Montrul & Sánchez-Walker 2013). The fact that Spanish L2 learners demonstrate less accuracy in oral than in written production might be due to the difficulty of accessing and assembling gender morphology in spontaneous speech, rather than to a representational deficit problem at the level of abstract syntactic formal features (cf. Bruhn de Garavito & White 2002; Lardiere 2006; Montrul, Foote & Perpiñán 2008). What may further account for the minor differences between these groups is the participants' language comfort when using Spanish and their amount of contact

with Spanish (see Carreira 2004; Potowski 2002; Valdés 1995). Heritage speakers acquire Spanish early in childhood primarily in a naturalistic setting (home) and find themselves more often in situations of spontaneous communication with other native or bilingual speakers who actively use the target language. Thus, their language comfort and motivation to speak Spanish might be higher than in Spanish L2 learners. Although oral communication activities are incorporated in Spanish language courses, the main focus lies on conveying knowledge about language or metalinguistic skills through written input and on fostering learners' writing skills (cf. Montrul, Foote & Perpiñán 2008). This is a crucial difference that sets apart Spanish L2 learners and heritage speakers "who are often schooled in the majority language, and many have little or no literacy experience in their L1 until later in life, when they enroll in classes to relearn the heritage language in a formal setting" (Montrul, Foote & Perpiñán 2008: 507).

Regarding the extra-linguistic variable input, Montrul, Foote and Perpiñán (2008) analyzed data from heritage speakers of Spanish. Their findings reveal that despite exposure to the language in early childhood, heritage speakers make similar kinds of errors compared to adult L2 learners. According to them, the non-target-like performance results from insufficient input. Similarly, Mueller-Gathercole and Thomas (2005) examined the effect of input and exposure in a study on the acquisition of the grammatical gender system in Welsh and reported that both input and exposure to Welsh at home and/or at school have an impact on target-like realizations. In particular, the factor of exposure at home aids learners to achieve native-like competence.

To summarize, recent studies on gender assignment and agreement are inconclusive as to whether language contact on an individual level affects gender accuracy in adult English-speaking L2 learners of Spanish compared to Spanish heritage speakers. Existing studies on bilingual child development indicate that both the amount and type of exposure play an important role in so far that a reduced amount of exposure has been argued to affect various linguistic domains (e.g. Unsworth 2013; Unsworth et al. 2014). Studies on late second language acquisition concur with these assumptions.

The present study examines whether the amount of exposure to the target language, i.e. Spanish, affects gender accuracy. In addition, the design of the study allows us to investigate whether Spanish L2 learners and heritage speakers of the same proficiency level show non-ultimate attainment and incomplete acquisition as discussed controversially in the literature. The purpose of this study is to compare knowledge of gender in Spanish L2 learners and Spanish heritage speakers who differ in their amount of exposure to languages at an individual level. This is an important variable, as regular language contact outside of a formal setting may lead to more gender accuracy.

5.   Research questions

The previously mentioned studies suggest that gender is a problematic grammatical category and a common source of errors for adult L2 learners and heritage speakers. Despite the ample research in this field, the question of whether the amount of language use may have an effect on the accuracy rate of gender agreement in Spanish L2 learners and Spanish heritage speakers is not yet fully explored. Thus, the present study addresses the following research questions:

1.   Will there be differences in the written comprehension and oral production of gender agreement in noun phrases between Spanish L2 learners and Spanish heritage speakers?
2.   If L2 learners and heritage speakers make gender agreement errors, will there be differences between the two groups with respect to the patterns of errors in terms of the domain of agreement (determiner, adjective)?
3.   Is there a relationship between the amount of exposure to Spanish and the accuracy rate of gender agreement?

In light of these three research questions the following predictions are made:

1.   There will be differences in the written comprehension and oral production of gender agreement in noun phrases between both groups. Spanish L2 learners will be more target-like than the heritage speakers in the written comprehension task, whereas the heritage speakers will be more target-like than the second language learners in the oral production task.
2.   There will be more errors in the domain of noun-adjective agreement than determiner-noun agreement within the DP. More precisely, it is predicted that Spanish L2 learners will produce more errors in the domain of noun-adjective agreement since it is acquired later than the congruency between the article and the noun due to its complexity.
3.   Assuming that the amount of exposure to Spanish affects the linguistic performance of the participants, it is expected that participants having regular contact with Spanish will be more accurate in gender agreement than those who are seldom in contact with the target language.

6. Methodology

6.1   Participants

There are two groups of participants involved in this experiment: a group of 27 advanced English-speaking learners of Spanish (late foreign language learners, 19 females and 8 males; mean age 19.4, range 18–21) and 27 Spanish heritage speakers (Spanish-English simultaneous bilinguals; 22 females, 5 males, average age 20.5, range 19–21) all of whom were undergraduate students at a US college in New York State and from middle class families. All participants were formally educated in English and began receiving formal instruction in Spanish during university/college education. In these Spanish-language classes (not specific heritage speaker classes), all participants were exposed to different Spanish varieties. The group of L2 learners began learning Spanish at the age of 13 or later (mean 13.78). They were all born and raised in the United States in English-speaking families. The participants reported that they have travelled to a Spanish-speaking country (range of stay from 1.5 weeks to 3 months). All participants were administered a language contact questionnaire which included variables such as the age of acquisition (henceforth AoA) of both languages, age, languages spoken, predominant country of residence and the type and frequency of using Spanish as the target language (TL). L2 learners of Spanish and heritage speakers are many times exposed to Spanish in varying proportions and in different contexts and activities. In order to assess the frequency of using Spanish, participants were asked to do a self-rating of their respective amount of language use by means of a 3-point Likert-type scale, which includes different contexts outside of the classroom and different activities (reading, watching TV, meeting Spanish-speaking friends etc.). In the questionnaire, participants were informed about the meaning of the three labels defining the frequency of Spanish usage i.e. 'regular use' (about 80 percent of the time), 'occasional use' (more than half of the time) and 'seldom use'. Ten heritage speakers and nine Spanish L2 learners declared that they regularly used Spanish; another ten heritage speakers and nine Spanish L2 learners reported that they used it occasionally and seven heritage speakers and nine Spanish L2 learners reported on seldom use of Spanish outside of the classroom. The L2 learners' self-assessment of Spanish was 3.9 (range 3–5, where 1 = low proficiency and 5 = native-like) and their self-assessment in English was 5. As expected, English is the dominant language in this group (see Table 2). To indicate potential language dominance in the groups, the mathematical difference between the proficiency (in percent) in Spanish and German was calculated (see Schmitz, Di Venanzio & Scherger 2016 for the degrees of language balance). 23 heritage speakers were raised by two Spanish-speaking parents. In the other four cases the heritage speakers had one parent born in a Latin American country and

Table 2. Overview of subjects' proficiency and language dominance

Group	Age	DELE score (Total: 50)	Oxford score (Total: 40)	% Difference	Self reported proficiency Spanish	Self reported proficiency English	Dominance
L2ers	Mean: 19.4 Range: 18–21	30.6% Mean: 40.3 Range:36–45	96.25% Mean: 38.5 Range: 37–40	−15.65	Mean: 3.9 Range: 1–5	Mean: 5 Range: 1–5	English
2L2ers	Mean: 20.5 Range: 19–21	81.4% Mean: 40.7 Range 36–47	94.5% Mean: 37.8 Range:35–38	−13.1	Mean: 4.39 Range: 1–5	Mean: 4.8 Range: 1–5	English

another English-speaking parent born in the US. All heritage speakers were born and raised in the United States. Most of them reported using Spanish as the main, in many cases exclusive mode of communication within the family (nuclear and extended). They self-rated as highly proficient in Spanish (self-rating of at least 4 on a scale of 1–5; mean = 4.39), while their mean self-assessed proficiency in English was 4.8 (range 4–5). Both groups took the same English and Spanish proficiency test. The English proficiency test consisted of a cloze passage with a blank for every seventh word (total 40 blanks), and participants were to select one of three choices for each blank. The Spanish proficiency test consisted of a cloze test and a reading comprehension task (with four multiple choice options for each blank) from a version of the *Diploma de Español como Lengua Extranjera* (DELE) which is the same test used in several other studies of L2 learners and heritage speakers (McCarthy 2007; Montrul 2005; Montrul Foote & Perpiñán 2008). The maximum score (both sections combined) was 50. The Cronbach's alpha reliability coefficient for both proficiency tests was .827, indicating reliability. The boxplot in Figure 1 illustrates the distribution of the L2 learners and heritage speakers in Spanish. The L2 learners' mean accuracy on the proficiency test ($M = 40.3$ $SD = 2.801$, range 36–45) did not differ from the mean score of the heritage speakers ($M = 40.7$ $SD = 2.386$, range 36–47) ($t[52] = 0.471$, $p = .640$). Since neither group differs in the proficiency level, the study investigates whether the accuracy of gender assignment and agreement between L2 learners and heritage speakers is also similar and if the participants' personal language contact outside of the classroom plays a role. Although the groups do not differ in their proficiency levels, it is important to note that differences between the two groups can be found in the way they use the language. The L2 learners mainly use the language in the classroom setting, abroad or when they meet with Spanish-speaking friends whereas heritage speakers use Spanish significantly more often outside of the classroom setting such as at home and during activities such as reading, watching TV, meeting Spanish-speaking peers and friends etc.
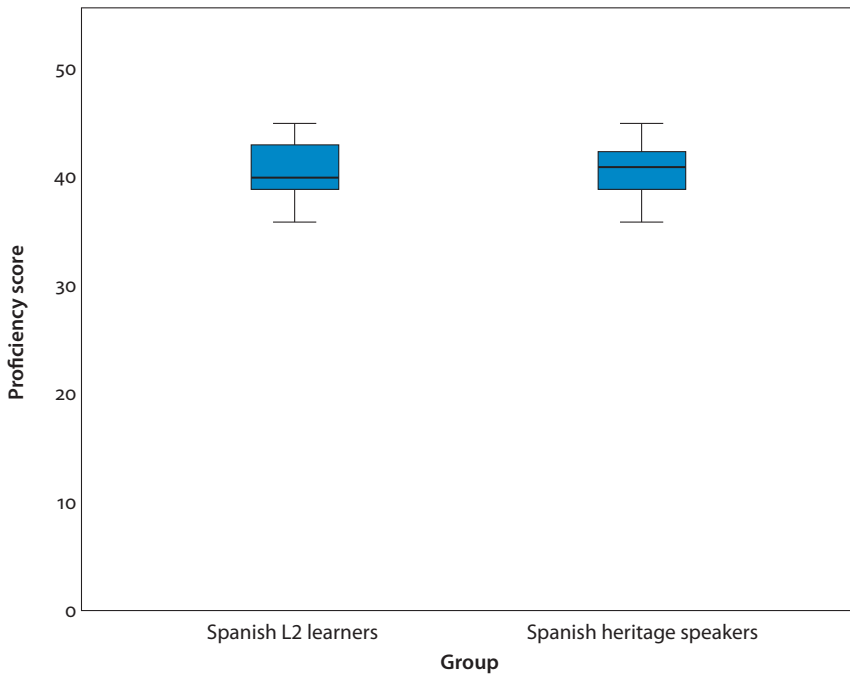
Figure 1. Distribution of proficiency scores (max. = 50) by group

6.2   Tasks and procedure

To test the predictions, the study followed a methodology used by Diebowski (2014) and Montrul, Foote & Perpiñán (2008). It employs two tasks: (1) a grammaticality judgement task (GJT) and (2) an oral production task (OPT). All participants were tested individually and in the presence of the researcher or an assistant. All tasks were untimed and the total testing time was between one and one and a half hour.

6.2.1   *Task 1: Grammatical judgment task (GJT)*

The first task (GJT) consists of a total of 40 items testing the Spanish heritage speakers' and L2 learners' knowledge of gender agreement in determiner-noun ($n = 20$) and noun-adjective combinations ($n = 20$) in the DP. The study included the following morphological classes of the modified noun:

a.   inanimate nouns gender-marked by -a and -o: *blusa*-FEM 'blouse', *libro*-MASC 'book'

b.   inanimate non-gender-marked nouns ending in -e: *leche*-FEM 'milk', *coche*-MASC 'car'

c.   inanimate deceptively marked nouns: *poema*-MASC 'poem'

The coded modifiers included attributive adjectives that are gender-marked by the endings -a and -o and the definite determiners *el* or *la*. All contexts contained nouns and adjectives which are familiar to both speaker groups. There was an equal number of ungrammatical and grammatical tokens, which were randomized. The test questions focused only on one token per sentence. The task included 20 fillers to mask the purpose of the study. Participants were asked to read each sentence and evaluate it as either acceptable or unacceptable. If they judged it to be unacceptable, they were instructed to modify the incorrect part. An example can be seen in (9).

(9) a. Juan busca *el foto de sus abuelos.
  "Juan looks for the.MASC* photo.FEM of his grandparents."
   b. Sergio no puede encontrar su mapa *amarilla.
  "Sergio cannot find his yellow.FEM* map.MASC."

In the Example (9a), the test question incorrectly includes the masculine determiner *el* 'the' used to modify the feminine noun *foto* 'photo', and in (9b) the feminine adjective *amarilla* 'yellow' to modify the masculine noun *mapa* 'map'. The aim was for the participants to correct the ungrammatical modifiers *el* and *amarilla* by the correct feminine determiner *la* and masculine adjective *amarillo.* Any correction other than these was disregarded.

6.2.2 Task 2: Oral picture description task (OPDT)

To investigate the knowledge of gender assignment and agreement in Spanish in oral production by L2 learners and heritage speakers, the study employed a picture description task (OPDT) similar to the one used by Montrul, Foote & Perpiñán (2008). The task consisted of 40 pictures. Half of them were test items and the other half were distractors. The stimuli were pictures of inanimate nouns which were equally distributed in terms of grammatical gender: 10 were masculine and 10 were feminine. Furthermore, the task included two classes of nouns based on their endings in -a vs. -o vs. -e (as described in Section 3.1). All pictures were presented one by one in a PowerPoint presentation and the participants were asked to describe what they saw on each picture by completing the carrier sentence *Veo un/una + noun + adjective* ("I see a adjective + noun") as exemplified in Figure 2.

The participants' responses were audio-recorded and later transcribed and coded for two categories: (1) the correct agreement between the determiners and noun and (2) the correct agreement between the noun and adjective. Correct responses received a score of one point and incorrect responses a score of zero.
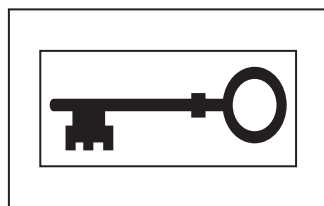
Target: *Veo una llave negra/ fina / bonita.*

"I see a black/ small/ beautiful key."

Figure 2. Example slide presentation and expected response
from the oral picture description task

7. Results

In this section, the results of the experiments and the statistical analyses are presented. Cronbach's alpha scores for all three tasks were greater than 0.08, indicating reliability. Furthermore, the Levene's test was used to assess whether the groups have equal variances. This test was not significant. Thus, the group variances can be treated as equal. All analyses used a significance level of $\alpha = 0.05$. For statistical purposes in both tasks, the study assessed the number of correct answers by both groups and reports on the effect of language contact at an individual level (range from using the target language seldomly to occasionally and regularly). In the following sections, the results will be discussed with respect to the linguistic behavior of the participants and the possible impact of the amount of language contact.

7.1 Task 1: Grammaticality judgment task (GJT)

Recall that in this task subjects were required to judge whether the sentence presented to them is acceptable or not. In the latter case, they were asked to provide a correction. Before analyzing the data statistically, all corrections given by the participants were checked. All target-like sentences that were judged as acceptable received a score of one and those that were judged as unacceptable a score of zero.

Figure 3 displays the ratio of correct agreement with the determiner produced by the group of Spanish heritage speakers and L2 learners, including the frequency of using the target language outside of the classroom. A one-way Analysis of Variance (ANOVA) was conducted on the agreement between determiner and noun, using the mean target-like realizations as the dependent variable and the subgroups of L2 learners and heritage speakers with their different amounts of language contact as the independent variable. The analyses showed a significant effect for the degree of language contact [$F(5, 48) = 3.383$, $p = 0.011$], with a strong

partial eta squared effect size of 0.261. As illustrated in Figure 3, there appear to be similarities and differences between the groups' accuracy rates for gender and the degree of language contact, so further statistical analyses were conducted. The heritage speakers ($M = 9.8$, $SD = 0.42$) and L2 learners ($M = 9.7$, $SD = 0.48$) with regular use of the target language had very comparable accuracy levels. The L2 learners who regularly used Spanish ($M = 9.3$, $SD = 0.50$) had the same accuracy rate as was observed for heritage speakers who used Spanish occasionally ($M = 9.7$, $SD = 0.48$). However, the Tukey post-hoc test revealed significant differences between the heritage speakers using Spanish regularly and Spanish L2 learners who seldomly used the target language ($p = 0.021$). A further significant difference was found between the Spanish L2 learners with regular use of the target language and the Spanish L2 learners with seldom use of the target language ($p = 0.034$).
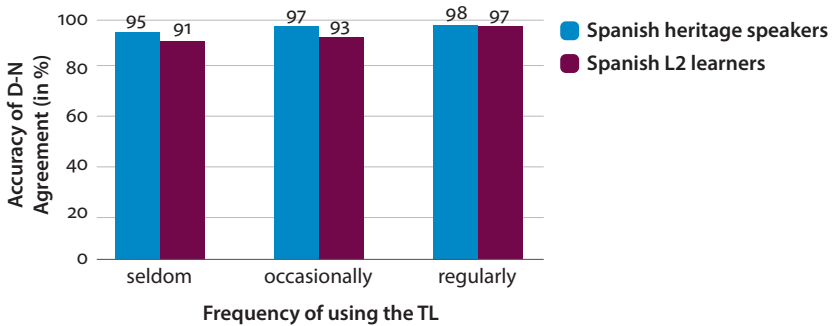
Figure 3. Accuracy of determiner-noun agreement in written judgments in a GJT in percent by group and frequency of using the TL outside of the classroom

Figure 4 shows the ratio of correct agreement between the noun and the adjective by group and frequency of use of the target language. A one-way Analysis of Variance (ANOVA) using the mean-target-like realizations as the dependent variable and the amount of language use as the independent variable showed a significant effect for the degree of language use [$F(5, 48) = 4.066$, $p = 0.004$], with a strong partial eta squared effect size of 0.298. The Tukey post-hoc test revealed, however, that significance is only found in the comparison between heritage speakers using Spanish regularly and L2 learners using Spanish infrequently ($p = 0.033$) as well as heritage speakers using Spanish occasionally and L2 learners using Spanish infrequently ($p = 0.003$). When comparing L2 learners who use Spanish occasionally and regularly with all three subgroups of heritage speakers based on their degree of language contact, it was found that they did not differ statistically from one another for the issue in question (see Figure 4).
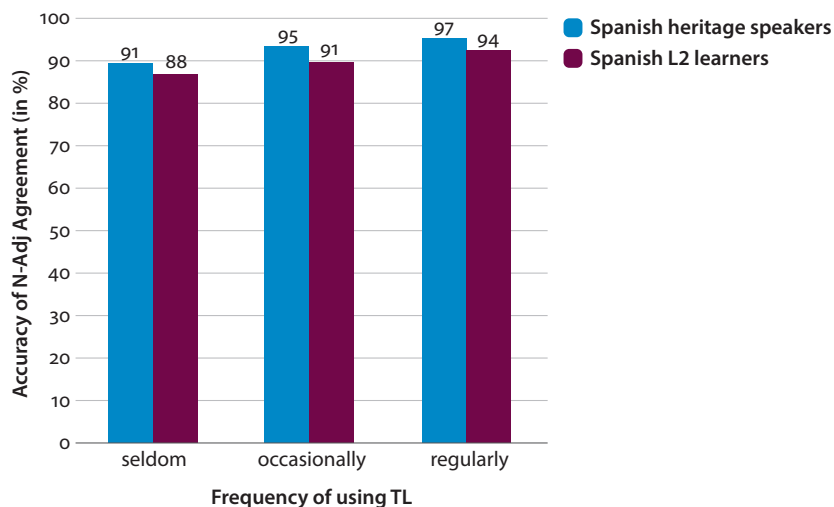
Figure 4. Accuracy of noun-adjective agreement in written judgments in a GJT in percent by group and frequency of using the TL outside of the classroom

7.2 Task 2: Oral picture description task (OPDT)

Let us turn to the analyses of gender accuracy in oral production, as measured by the Picture Description Task. Figure 5 presents the results for the ratio of correct agreement with the determiner by group and use of the target language. As can be observed, the Spanish L2 learners with rare exposure to the target language exhibit the lowest gender accuracy rate (89%). Looking more closely at the accuracy rates in the L2 learners' subgroups based on the degree of language contact, a clear trend is evident. The more contact the L2 learners have with the target language, the more target-like is their performance on gender agreement. To determine if any apparent differences between the groups were significant, a one-way ANOVA was conducted. Using the mean-target-like realizations as the dependent variable and the frequency of use of Spanish as the independent variable, a significant effect for the degree of language contact [$F(5, 48) = 7.691$, $p = 0.000$], with a strong partial eta squared effect size of 0.445 was found. The Tukey post-hoc test revealed significant differences between the group of L2 learners having seldom contact with the target language and all the HS subgroups. The same effect holds for the group of L2 learners having seldom contact with the target language and the group of L2 learners having regular contact with the target language. However, no statistically significant difference between the group of L2 learners having seldom contact with
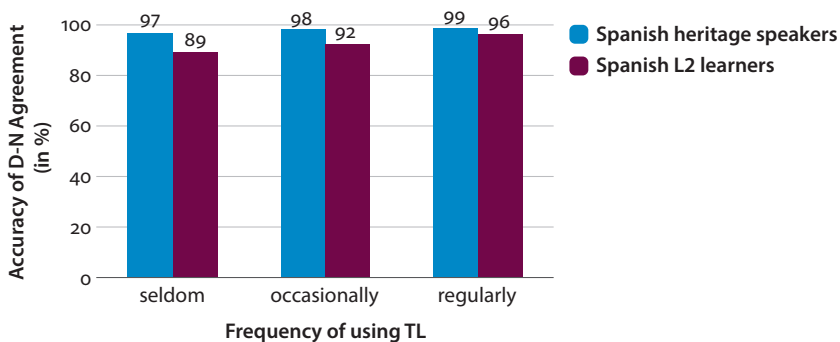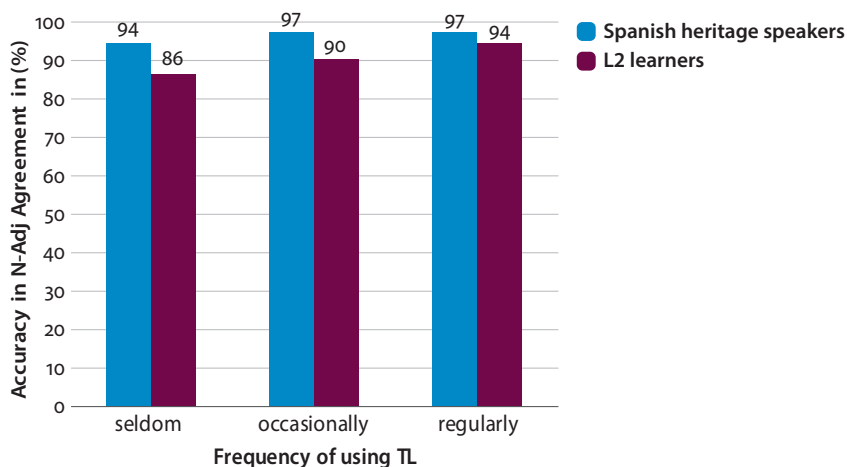
Figure 5. Accuracy of determiner-noun agreement in oral production in percent by group and frequency of using the TL outside of the classroom

the target language and those L2 learners who have occasional contact with the target language could be found.

Regarding noun-adjective agreement in the oral production task, the trend toward a higher frequency of language use and higher gender accuracy rates is also evident in the group of Spanish L2 learners. L2 learners with regular contact with the target language outperform those with less language contact. This trend does not apply to the subgroups of heritage speakers. The results of the oral production task show that heritage speakers having occasional language contact and heritage speakers having regular language contact produce identical accuracy rates regarding agreement between the noun and the adjective (97%). A one-way ANOVA using the mean-target-like realizations as the dependent variable and the amount of language contact as the independent variable showed a significant effect for the degree of language contact [$F(5, 48) = 4.421$, $p = 0.002$], with a strong partial eta squared effect size of 0.315. The Tukey post-hoc tests comparing the subgroup of L2 learners having seldom language contact to all other subgroups, except for the L2 learners having occasional language contact, were all significant. It was found that the performance of L2 learners having seldom language contact differs statistically significantly from that of L2 learners having regular language contact ($p = 0.041$). Likewise, the performance of L2 learners having seldom language contact differs statistically significantly from the one of heritage speakers having occasional ($p = 0.007$) and regular language contact ($p = 0.003$).

Figure 6. Accuracy of noun-adjective agreement in oral production in percent by group and frequency of using the TL outside of the classroom

8.   Discussion and conclusion

The current study investigated the overall accuracy of gender assignment and agreement in Spanish L2 learners and Spanish heritage speakers in the United States with different frequency rates of using the target language. Under the assumption that regular use of Spanish exerts a significant influence in the accuracy on gender marking, the study investigated the potential role of the amount of language use at the morphosyntactic level in Spanish L2 learners and heritage speakers. Based on the results, the study reveals that Spanish L2 learners and heritage speakers showed a high overall accuracy.

The aim of the study was to answer the research questions posed earlier in the chapter:

1.   Will there be differences in the written comprehension and oral production of gender agreement in noun phrases between Spanish L2 learners and Spanish heritage speakers?

The data show only minimal differences between the written comprehension and oral production of gender agreement in noun phrases between Spanish L2 learners and Spanish heritage speakers. The performance of the Spanish L2 learners suggests that they produce fewer gender agreement inaccuracies overall in the written comprehension than in oral production. According to

the L2 acquisition literature reviewed, L2 learners produce fewer gender agreement inaccuracies overall in written comprehension than in oral production, whereas heritage speakers have little or no literacy experience in their heritage language until they enroll in classes to relearn the heritage language in a formal setting (Montrul, Foote & Perpiñán 2008; Alarcón 2006; Bruhn de Garavito & White 2002; Lardiere 2006). In line with Kupisch, Akpinar & Stöhr (2013), but against the findings by Montrul, Foote & Perpiñán (2008) and Alarcón (2006), the results of the Spanish L2 learners and heritage speakers in this study could not show any statistically significant differences with regard to gender accuracy between L2 learners and heritage speakers in the written versus the oral task.

2. If L2 learners and heritage speakers make gender agreement errors, will there be differences between the two groups with respect to the patterns of errors in terms of the domain of agreement (determiner, adjective)?

    The data lend some support to the conclusion that overall the Spanish L2 learners showed a greater ratio of errors in gender agreement between noun and adjective than between article and noun (see Figures 3–6). This finding concurs with previous research by Finnemann (1992), Fernández-García (1999) and Bruhn de Garavito & White (2002) who observed lower accuracy rates for adjectives than for determiners, suggesting that adjective agreement is acquired later than agreement with articles. However, observing the individual groups of speakers, there were some differences: The Spanish L2 learner group tended to produce a slightly higher ratio of gender-agreement inaccuracies between noun and adjective than between determiner and noun. This can be attributed to the fact that noun-adjective agreement is linguistically a more demanding task for L2 learners than article-noun agreement. The heritage speaker group, however, supplied about the same ratio of gender-agreement inaccuracies in each of these categories.

3. Is there a relationship between language contact and the accuracy rate on gender agreement?

    The beneficial effect of the amount of language use could not be found for the linguistic performance of the Spanish L2 learner groups and heritage speaker groups across the three levels of language contact (i.e. seldom vs. occasional vs. regular use of Spanish). In addition, there was no effect of language contact on gender accuracy among the different heritage speaker subgroups. Nevertheless, an effect was indeed evident for heritage speakers and L2 learners with regular or occasional contact with the target language as they outperformed those L2 learners who had seldom contact with Spanish. This finding concurs with previous research by Dewaele and Véronique (2001: 292) who concluded that the acquisition of gender assignment and agreement by foreign language learners

is less affected by the amount of teaching than by the amount of authentic communication with native speakers outside the classroom. In the same vein, Flege and Liu (2001: 257) concluded that a higher amount of native speaker input will noticeably improve learners' performance in the foreign language.

In conclusion, the present study has found an effect of language contact on higher gender accuracy rates by Spanish L2 learners who have regular or occasional contact as opposed to those L2 learners who have seldom contact with the target language. In the case of Spanish heritage speakers, the study could not find any effect between the amount of language use and gender accuracy. The results indicated that heritage speakers, who use their heritage language rather frequently or infrequently, perform almost at ceiling in all the tasks. These results allow for the conclusion that gender assignment and agreement rules are seemingly not very vulnerable with regard to language attrition within the group of heritage speakers involved in this study. However, we must keep in mind that all heritage speakers involved in this study attended classes in Spanish which might also be a relevant factor for the high gender accuracy rates. This does not preclude the possibility that an effect of the amount of language exposure may be found when testing more items (especially infrequent lexical items or nonce words) and other syntactic contexts within larger groups of speakers.

Moreover, the data from this study shows that both the Spanish L2 learners and the heritage speakers do not have grammatical deficits in gender assignment or agreement. Contrary to the claim of incomplete acquisition put forward by various scholars, among them Montrul, Foote & Perpiñán (2008) and Polinsky (2008), this study supports the view that heritage speakers and L2 learners can successfully acquire gender agreement. The high accuracy rates (see Figures 3–6) sustain this argument. However, we should bear in mind that heritage speakers form a very heterogenous group since they differ in terms of age of acquisition, language input and activation of the heritage language (Kanno et al. 2008; Kondo-Brown 2005; Valdés 1995).

Clearly, a limitation of this study is the low number of participants in the Spanish L2 learner and heritage speaker subgroups. Other scholars might question the generalizability of the obtained results. On this basis, the results should be considered as indicative and not conclusive since larger groups, more questions on the language contact profile regarding the type of language instruction received and a combination of online and offline tasks could provide support for the argument that gender accuracy in the target language depends on the amount of language contact.

From the results reported here, several conclusions for teaching L2 learners and heritage speakers can be drawn. From the perspective of research on L2 acquisition (SLA), the category gender is a difficult grammatical phenomenon for L2 learners

despite many years of instruction in a formal classroom-setting. It is important that grammatical structures are taught to develop communicative competence in the target-language. The approach of communicative grammar teaching equips the learners with strategies and activities to develop communicative competence marked by fluency as well as grammatical accuracy. With respect to heritage speakers, whose profiles differ from Spanish L2 learners, language instructors have to be aware of the special needs of these speakers. A Spanish class in which heritage speakers and L2 students are taught together provides many challenges and often does not address the linguistic needs of heritage speakers. Therefore, it is important to develop pedagogies and policies suitable for teaching Spanish to heritage speakers (cf. Valdés 2001: 12).

Acknowledgements

References

Alarcón, I. V. 2006. *The Second Language Acquisition of Spanish Gender Agreement: The Effects of Linguistic Variables on Accuracy*. Munich: Lincom.

Alarcón, I. 2008. The language of language: Core concepts in linguistic analysis. *The Modern Language Journal* 92(3): 477–478. https://doi.org/10.1111/j.1540-4781.2008.00759_4.x

Appel, R. & Muysken, P. 2005. *Language Contact and Bilingualism*. Amsterdam: Amsterdam University Press.

Bloomfield, L. 1935. *Language*. London: Allen and Unwin.

Bolonyai, A. 2007. (In)vulnerable agreement in incomplete bilingual L1 learners. *International Journal of Bilingualism* 11: 3–21. https://doi.org/10.1177/13670069070110010201

Bosque, I. 2007. *Las categorías gramaticales*, 7th edn. Madrid: Sintesis.

Bruhn de Garavito, J. & White, L. 2002. The second language acquisition of Spanish DPs: The status of grammatical features. In *The Acquisition of Spanish Morphosyntax: The L1/L2 Connection*, A. T. Pérez-Leroux & J. Muñoz Liceras (eds), 153–178. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-010-0291-2_6

Cain, J., Weber-Olsen, M. & Smith, R. 1987. Acquisition strategies in a first and second language: Are they the same? *Journal of Child Language* 14(2): 333–352. https://doi.org/10.1017/S0305000900012964

Cantone, K. F. & Müller, N. 2008. *Un Nase* or *una Nase*? What gender marking within switched DPs reveals about the architecture of the bilingual language faculty. *Lingua* 118(6): 810–826. https://doi.org/10.1016/j.lingua.2007.05.007

Carreira, M. 2004. Seeking explanatory adequacy: A dual approach to understanding the term "heritage language learner." *Heritage Language Journal* 2(1). <http://www.heritagelanguages.org> (20 May 2008).

Carroll, S. 1999. Input and SLA: Adult's sensitivity to different sorts of cues to French gender. *Language Learning* 49(1): 37–92. https://doi.org/10.1111/1467-9922.00070

Chini, M. 1995. Un aspect du syntagme nominal en italien L2: le genre. *AILE Acquisition et Interaction en Langue Etagere* 5: 115–142.

Corbett, G. G. 1991. *Gender*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139166119

Cruzan, A. 2003. *Gender Shifts in the History of English*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511486913

Dewaele, J. M. & Véronique, D. 2001. Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study. *Bilingualism: Language and Cognition* 4: 275–297. https://doi.org/10.1017/S136672890100044X

Diebowski, J. 2014. On the effects of proficiency and input in the acquisition of grammatical gender in Spanish. *Journal of Teaching and Education* 3(1): 313–322. <http://www.universitypublications.net/jte/0301/pdf/R3ME339.pdf>

Eichler, N., Jansen, V. & Müller, N. 2013. Gender in French-German, Italian-German and Italian-French children. *International Journal of Bilingualism* 17(5): 550–572. https://doi.org/10.1177/1367006911435719

Fernández-García, M. 1999. Patterns of gender agreement in the speech of second language learners. In *Advances in Hispanic Linguistics: Papers from the 2nd Hispanic Linguistics Symposium*, J. Gutiérrez-Rexach & F. Martínez-Gil (eds), 3–15. Somerville MA: Cascadilla Press.

Filppula, M., Klemola, J. & Paulasto, H. (eds). 2009. *Vernacular Universals and Language Contacts. Evidence from Varieties of English and Beyond*. London: Routledge.

Finnemann, M. D. 1992. Learning agreement in the noun phrase: The strategies of three first-year Spanish students. *International Review of Applied Linguistics in Language Teaching* 30(2): 121–136. https://doi.org/10.1515/iral.1992.30.2.121

Flege, J. E., & Liu, S. 2001. The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition* 23: 527–552.

Franceschina, F. 2001. Against an L2 morphological deficit as an explanation for the differences between native and non-native grammars. *EUROSLA Yearbook* 1: 143–158. https://doi.org/10.1075/eurosla.1.12fra

Gess, R. & Herschensohn, J. 2001. Shifting the DP parameter: A study of Anglophone French L2ers. In *Romance Syntax, Semantics and their L2 Acquisition* [Current Issues in Linguistic Theory 216], J. Camps & C. R. Wiltshire (eds), 105–119. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.216.10ges

Green, J. N. 1988. Spanish. In *The Romance Languages*, M. Harris & N. Vincent (eds), 79–130. Oxford: OUP.

Hawkins, R. & Franceschina, F. 2003. Explaining the acquisition and non-acquisition of determiner-noun gender concord in French and Spanish. In *The Acquisition of French in Different Contexts* [Language Acquisition and Language Disorders 32], P. Prévost & J. Paradis (eds), 175–205. Amsterdam: John Benjamins.

Herschensohn, J. & Arteaga, D. 2015. Parameters, processing and feature reasssembly in the L2 French determiner phrase. In *The Acquisition of French in Multilingual Contexts*, P. Guijarro-Fuentes, K. Schmitz & N. Müller (eds), 215–233. Bristol: Multilingual Matters. https://doi.org/10.21832/9781783094530-011

Hockett, C. F. 1958. *A Course in Modern Linguistics*. New York NY: The Macmillan Company. https://doi.org/10.1111/j.1467-1770.1958.tb00870.x

Jaensch, C. 2009. L3 enhanced feature sensitivity as a result of higher proficiency in the L2. In *Third Language Acquisition and Universal Grammar*, Y.-K. Leung (ed), 115–143. Bristol: Multilingual Matters.

Jarvis, S. & Pavlenko, A. 2008. *Cross-linguistic Influence in Language and Cognition*. New York NY: Routledge.

Kanno, K., Hasegawa, T., Ikeda, K., Ito, Y. & Long, M. H. 2008. Prior language-learning experience and variation in the linguistic profiles of advanced English-speaking learners of Japanese. In *Heritage Language Education: A New Field Emerging*, D. M. Brinton, O. Kagan & S. Bauckus (eds), 321–333. New York NY: Routledge.

Kolehmainen, L., Meriläinen, L. & Riionheimo, H. 2014. Interlingual reduction: Evidence from language contacts, translation and second language acquisition. In *Language Contacts at the Crossroads of Disciplines*, H. Paulasto, L. Meriläinen, H. Riionheimo & M. Kok (eds), 3–32. Newcastle upon Tyne: Cambridge Scholars.

Kondo-Brown, K. 2005. Differences in language skills: Heritage language learner subgroups and foreign language learners. *The Modern Language Journal* 89(4): 563–581. https://doi.org/10.1111/j.1540-4781.2005.00330.x

Kupisch, T., Akpınar, D. & Stöhr, A. 2013. Gender assignment and gender agreement in adult bilinguals and second language learners of French. *Linguistic Approaches to Bilingualism* 3(2): 150–179. https://doi.org/10.1075/lab.3.2.02kup

Lardiere, D. 2006. *The Fossilized Steady State in Second Language Acquisition*. Mahwah NJ: Lawrence Erlbaum Associates.

Lipski, J. M. 1993. Creoloid phenomena in the Spanish of transitional bilinguals. In *Spanish in the United States: Linguistic Contact and Diversity*, A. Roca & J. Lipski (eds), 155–182. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110804973.155

Lipski, J. M. 1996. Patterns of pronominal evolution in Cuban-American bilinguals. In *Spanish in Contact*, A. Roca & J. B. Jensen (eds), 159–186. Somerville MA: Cascadilla Press.

Martínez-Gibson, E. A. 2011. A comparative study on gender agreement errors in the spoken Spanish of heritage speakers and second language learners. *Porta Linguarum* 15: 177–193.

McCarthy, C. 2007. Morphological Variability in Second Language Spanish. PhD dissertation, McGill University, Montreal, Canada.

Montrul, S. 2002. Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition* 5(1): 39–68. https://doi.org/10.1017/S1366728902000135

Montrul, S. 2004a. Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and Cognition* 7(2): 125–142. https://doi.org/10.1017/S1366728904001464

Montrul, S. 2004b. *The Acquisition of Spanish: Morphosyntactic Development in Monolingual and Bilingual L1 Acquisition and Adult L2 Acquisition* [Language Acquisition and Language Disorders, 37]. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.37

Montrul, S. 2005. Second language acquisition and first language loss in adult early bilinguals: Exploring some differences and similarities. *Second Language Research* 21: 1–51. https://doi.org/10.1191/0267658305sr2470a

Montrul, S. 2006a. Incomplete acquisition as a feature of bilingual and L2 grammars. In *Inquiries in Linguistic Development: In Honor of Lydia White*, R. Slabakova, S. Montrul & P. Prévost (eds), 335–359. Amsterdam: John Benjamins. https://doi.org/10.1075/z.133.18mon

Montrul, S. 2006b. Incomplete acquisition in bilingualism as an instance of language change. In *L2 Acquisition and Creole Genesis* [Language Acquisition and Language Disorders 42], C. Lefebvre, L. White & C. Jourdan (eds), 379–400. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.42.22mon

Montrul, S., Foote, R. & Perpiñán, S. 2008. Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning* 58(3): 503–553. https://doi.org/10.1111/j.1467-9922.2008.00449.x

Montrul, S. & Sánchez-Walker, N. 2013. Differential object marking in child and adult Spanish heritage speakers. *Language Acquisition* 20(2): 109–132. https://doi.org/10.1080/10489223.2013.766741

Mueller-Gathercole, V. & Thomas, E. M. 2005. Minority language survival: Input factors influencing the acquisition of Welsh. In *Proceedings of the 4th International Symposium on Bilingualism*, J. Cohen, K. T. McAlister, K. Rolstad & J. MacSwan (eds), 852–874. Somerville MA: Cascadilla Press.

Pascual y Cabo, D. & Rothman, J. 2012. The (il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied Linguistics* 33(4), 450–455. https://doi.org/10.1093/applin/ams037

Picallo, M. C. 1991. Nominals and nominalization in Catalan. *Probus* 3: 279–316. https://doi.org/10.1515/prbs.1991.3.3.279

Polinsky, M. 2007. Incomplete acquisition: American Russian. *Journal of Slavic Linguistics* 14: 191–262.

Polinsky, M. 2008. Russian gender under incomplete acquisition. *Heritage Language Journal* 6(1): 40–71.

Potowski, K. 2002. Experiences of Spanish heritage speakers in university foreign language courses and implications for teacher training. *ADFL Bulletin* 33(3): 35–42. https://doi.org/10.1632/adfl.33.3.35

Ramírez, D. B. 2003. L1 regression in Spanish-English bilinguals. In *Selected Proceedings of the First Workshop on Spanish Sociolinguistics*, L. Sayahi (ed), 89–95. Somerville MA: Cascadilla Proceedings Project.

Roca, I. M. 1989. The organization of grammatical gender. *Transactions of the Philological Society* 87(1): 1–3. https://doi.org/10.1111/j.1467-968X.1989.tb00617.x

Rothman, J. & Treffers-Daller, J. 2014. A prolegomenon to the construct of the native speaker: Heritage speaker bilinguals are natives too! *Applied Linguistics* 35(1): 93–98. https://doi.org/10.1093/applin/amt049

Sagarra, N. & Herschensohn, J. 2011. Asymmetries in gender and number agreement processing in late bilinguals. In *Selected Proceedings of the 13th Hispanic Linguistics Symposium*, L. A. Ortiz-López (ed.), 169–177. Somerville MA: Cascadilla Press.

Schmitz, K., Di Venanzio, L. & Scherger, A.-L. 2016. Null and overt subjects in Italian and Spanish heritage speakers in Germany. *Lingua* 180: 101–123. https://doi.org/10.1016/j.lingua.2016.04.004

Silva-Corvalán, C. 1994. *Language Contact and Change*. Oxford: OUP.

Silva-Corvalán, C. 2003. Linguistic consequences of reduced input in bilingual first language acquisition. In *Linguistic Theory and Language Development in Hispanic Languages*, S. Montrul & F. Ordóñez (eds), 375–397. Somerville MA: Cascadilla Press.

Silva-Corvalán, C. & Treffers-Daller, J. (eds). 2016. *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*. Cambridge: CUP. https://doi.org/10.1017/CBO9781107375345

Teschner, R. V. & Russell, W. M. 1984. The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics* 1: 115–132.

Thomason, G. S. & Kaufman, T. 1991. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley CA: University of California Press.

Toribio, J. 2001. On the emergence of bilingual code-switching competence. *Bilingualism: Language and Cognition* 4: 203–231. https://doi.org/10.1017/S1366728901000414

Tucker, G. R., Lambert, W. E. & Rigault, A. A. 1977. *The French Speaker's Skill with Grammatical Gender: An Example of Rule-Governed Behavior*. The Hague: Mouton.

Unsworth, S. 2013. Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition* 16: 86–110. https://doi.org/10.1017/S1366728912000284

Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A. & Tsimpli, I. 2014. On the role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics* 35: 765–805. https://doi.org/10.1017/S0142716412000574

Valdés, G. 1995. The teaching of minority languages as "foreign" languages: Pedagogical and theoretical challenges. *The Modern Language Journal* 79(3): 299–328. https://doi.org/10.1111/j.1540-4781.1995.tb01106.x

Valdés, G. 2001. Heritage language students: profiles and possibilities. In *Heritage Languages in America: Preserving a National Resource*, J. Peyton, J. Ranard & S. McGinnis (eds), 1–42. McHenry IL: The Center for Applied Linguistics and Delta Systems.

Weinreich, U. 1953. *Languages in Contact*. The Hague: Mouton.

White, L. 2003. *Second Language Acquisition and Universal Grammar*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511815065

Zagona, K. 2002. *The Syntax of Spanish*. Cambridge: CUP.

Vocabulary development in the heritage languages Russian and Turkish between ages 6 and 10

How do parental input and socio-economic status account for differences within and between the cohorts?

Elke G. Montanari[i], Roman Abel[ii], Lilia Tschudinovski[i] and Barbara Graßer[i]

[i]University Hildesheim / [ii]University of Kassel

How does vocabulary in the heritage language develop? Does the social environment of the community have an influence? This chapter presents empirical results regarding the development of expressive and receptive vocabulary in the heritage language and analyzes the effects of amount of exposure, use, socio-economic status, dominance, and community on the acquisition of vocabulary in the heritage language. A Russian-German and a Turkish-German speaking sample are compared: 211 children at the age of 6–10 years were tested with a standardized picture naming task in a cross-sectional design. The results show a good receptive mastery and a limited expressive command of vocabulary with large individual differences, and only a slight development in the timespan of four years. Between the communities we find systematic variation, which we attribute to social and pragmatic differences with moderator and mediator effects. Possible limitations of the results are discussed with respect to cross-linguistic test effects.

Keywords: vocabulary development, heritage Russian, heritage Turkish, amount of exposure, socio-economic status, dominance, community

1. Introduction

The UN Convention on the Rights of a Child underlines that the signing States Parties agree that education should aim to develop respect for the child's parents, as well as his or her own cultural identity, language and values (Art. 29, d). Children belonging to minority groups should not be denied the use of his or her

own language (Art. 30) (UNCRC 1989). In many parts of Germany, linguistic diversity, which implies the use of heritage languages for many purposes, in public as well as in private areas, is common (Siemund, Gogolin, Schulz & Davydova 2013). In the education system, children are sometimes explicitly forbidden from using their heritage language, as, in many cases, it is regarded as a hindrance for the acquisition of the German language (Montanari 2017). The focus at school is still on monolingual education in the majority language (Gogolin et al. 2011). Heritage language support is offered in some schools, but the bulk of the responsibility for heritage language education lies with the families or parents' networks.

2. State of research and research questions

Mastery of the heritage language has been found to be different from learners' mastery of a foreign language and monolinguals' mastery regarding grammatical and lexical competence; heritage language speakers have been found to apply particular strategies to cover lexical retrieval problems and they behave differently than monolinguals in association tasks (Isurin & Ivanova-Sullivan 2008; Kondo-Brown 2005; Shin Kim 2013) and show differences in word order due to contact with the majority language (Brehmer & Usanova 2015). There is some controversy when it comes to arguing for or against (in-)completeness of acquisition (Montrul 2008). The reasons to assume a specific acquisition path for the heritage language are mainly based on the following arguments: the age of onset is not always in the first year of life, the input, i.e. the language of the parents, often shows language contact and attrition influences, and typically the input in the heritage language derives from a smaller variety of speakers than in other acquisition constellations (Kupisch, Belikova, Özçelik, Stangen & White 2017; Putnam & Sánchez 2013). Past studies have demonstrated that socio-economic status and educational experience influence interaction within the family and bilingual education (Hoff & Tian 2005). Socio-economic status (Smithson et al. 2014) has an impact on the parents' level of education. The level of education influences whether or not parents consider maintaining the heritage language to be important; high levels of education are often associated with a stronger wish to maintain the heritage language (Brizić 2007; Lambert & Taylor 1996). Input, amount of exposure and literacy play important roles in maintaining the heritage language (Pfaff 1994; Rehbein, Herkenrath & Karakoç 2009; Sürig, Şimşek, Schroeder & Boneß 2016).

For the aforementioned reasons, it is an analytical necessity to consider the linguistic environment. The linguistic environment in migration and heritage language contexts differ regarding historical, social and political constraints, the moment of migration, the length of residence in the new country and the organization in communities. *Community* is a vague, yet, for first access attempts, expedient

concept of one of the social environments in situations created by migration. We understand communities as structures of specific cultural and linguistic practices (Scheele, Lesemann & Mayo 2010). How close-knit a network is exerts an influence on the linguistic means used within and outside the network, which vary depending on the extent to which the individual is integrated into the network (Milroy 1980). Social networks play a supporting role in migration situations (Haug 2000; Schweizer 1996; Gurak & Caces 1992) and influence key decision-making processes (Aichinger 1997). Highly integrative networks result in language maintenance (Vetter 1997: 137). The duration of stay and the generation status, such as immigrant children versus children *of* immigrants, influence language choice, dominance, proficiency and literacy (Andrews 1999).

Community membership could (a) differ systematically with regard to the input situation and other language maintenance factors and could therefore constitute a *mediator* relation within the statistical analysis, or (b) *moderate* the impact of influential factors – i.e. strengthen or weaken their connection to the acquisition of vocabulary in the heritage language, or even change its direction, or (c) be included in the statistical analysis as a further *independent* variable, which, under the control of the main and interaction effects of stated factors, explains other incremental proportions of variance. The moderator hypothesis states that affiliation to the community moderates the effect of factors on language acquisition. This would be the case, for example, if one factor in community A has a different effect than in community B. The mediator hypothesis states that factor A has an effect on factor B, and B has an effect on heritage language acquisition. Therefore, B is the mediator. We expect the impact of the community on heritage language acquisition to be mediated by community characteristics such as language practices, education levels and socio-economic status. Furthermore, we expect the impact of these characteristics on heritage language acquisition to be moderated by the community. In other words, our hypothesis is that the degree and even the direction of the impact of certain factors should differ depending on the community the child belongs to.

We focus on vocabulary, as vocabulary mastery is a strong predictor for grammatical and pragmatic language development (Lee 2011) and international research has been conducted in different multilingual contexts (Bialystok, Luk, Peets & Yang 2010; Hoff et al. 2012; Pearson, Fernández & Oller 1993). To answer the question of whether and how input and vocabulary acquisition are influenced by the community, children from two large communities in Germany – the Russian and the Turkish community – are compared. Are there systematic differences between the two communities with regard to the linguistic situation in the family environment? What are the differences in receptive and productive vocabulary knowledge in German and in the heritage language among Turkish and Russian children from grades 1–4? What impact does input have on the heritage language? How do the parents' qualification levels affect children's vocabulary size in the heritage language?

3. The study

3.1   Sample and method

3.1.1   *Sample*

The sample is composed of two groups. We tested children from households in which at least one parent is a native speaker of Russian (*n* = 113, female = 56) or Turkish (*n* = 98, female = 52). The environment language is Standard German. The language(s) spoken at home are Russian (RU) or Turkish (TR) and/or German and eventually other languages, such as Kurdish or Ukrainian. We controlled for participants' socio-economic status. We did not control for bi- versus multilinguals.

To recruit participants, we contacted primary schools and parent associations, predominantly in Northern Germany. If a school agreed to participate, we tested all the multilingual Russian and Turkish children available, namely all the multilingual children that the school or the association assigned to us, in a kind of fishing net approach ('you take what you can get'). It has to be taken into consideration that the schools which chose to participate may have chosen to do so because they encourage multilingualism in their multilingual students (rather than focus purely on the majority language), which might influence the results.

The Russian-speaking children in the sample were aged between 6;0 and 10;11 years (*M* = 8;6, *SD* = 1;3). A total of 1.8% of the pupils attended kindergarten, 17.7% were from the first grade, 30.1% from the second, 24.8% from the third and 25.7% from the fourth. In the Russian group, 88.5% of the children were born in Germany. 81.1% had been exposed to German since birth and therefore had an age of onset of zero. The age of the children in the sample of Turkish speakers ranged from 6;3 to 10;10 years (*M* = 8;10, *SD* = 1;2). 13.3% were first grade pupils, 28.6% second grade, 34.7% third grade and 23.5% fourth grade. All of the children in the Turkish-speaking group were born in Germany. 45.6% of the Turkish-speaking children had been exposed to German since birth.

Regarding parental proficiency in the heritage language, most of the Russian parents rated their language skills, and hence the level of input for the child, to be at a native-speaker level or just below that level: 97.4% of the mothers and 82.6% of the fathers stated that they had a good to very good proficiency in Russian (Common European Framework of Reference for Languages CEFR levels B2 or C). The results generated in the Turkish group are similar: 87.7% of the Turkish mothers and 94% of the fathers were of the opinion that they had a good to very good proficiency in Turkish (CEFR levels B2 or C).

With respect to socio-economic status and educational level, the two communities did not differ significantly with regard to income, measured on the ordinal scale. On average, the families earned just over € 2,000 net per month. When analyzing the highest qualifications, the following picture emerges: 47.7% of the

Russian mothers had a university degree compared to 17.9% of the Turkish mothers. 16.3% of the Russian and 26.8% of the Turkish mothers had a higher education entrance certificate. 36% of the Russian mothers had a secondary school education compared to 48.2% of the Turkish mothers. In the case of the fathers, 46.3% of the Russian- and 25% of the Turkish-speaking groups had a university degree. 17.1% of the Russian-speaking fathers had gained a higher education entrance certificate, and 23.1% of the Turkish-speaking fathers. 36.6% of the Russian-speaking fathers had a secondary school education compared to 38.4% of the Turkish-speaking fathers. Parents without a school-leaving certificate were only found in the Turkish group – namely 7.1% of the mothers and 13.5% of the fathers. In the Russian group, 66.3% of the mothers and 69.9% of the fathers obtained their highest education level in Russia. In the Turkish group, 35.7% of the mothers and 50% of the fathers passed their last exam in Turkey. In the Russian group, considerably more parents obtained the highest degree of education in their native country than in the Turkish group. In addition, the levels of education of the mothers and fathers in the Russian-speaking group correlate strongly $r = .473^{**}$, while no such correlation is observed in the Turkish-speaking group.

We parameterized the learned and practiced professions of the parents of the participants with the International Socio-Economic *Index* of Occupational Status ISEI (Ganzeboom & Treiman 2003). For all parents, the social status of what they learned was higher than the profession they ended up practicing. Turkish and Russian parents significantly differ in terms of the status of the parents' learned and practiced profession: the learned and practiced professions of the Russian-speaking parents yield higher index values on the median than for the Turkish parents: mothers – learned: 55 (RU) vs. 32 (TR); fathers – learned: 49 (RU) vs. 34 (TR); mothers – practiced: 38 (RU) vs. 0 (TR); fathers – practiced: 39.5 (RU) vs. 32.5 (TR).[1]

3.1.2   *Method*

We used a cross-sectional design, and combined a standardized vocabulary and word-finding test for 6- to 10-year-olds, the vocabulary and word-finding test *Wortschatz- und Wortfindungstest für 6- bis 10-Jährige (WWT 6 10)* (Glück 2011), with a parental questionnaire and a pupil interview. This offers the possibility to have data for statistical analysis for the heritage language as well as the majority language and to combine the results.

For the WWT 6–10, a German standardized and a Turkish version were already available. The test was adapted for Russian, using translations and peer ratings by three adult native speakers of Russian. In the long version used in our project 95 words (26 nouns, 23 verbs, 23 adjectives, 23 generic terms=nouns) were used.

---

1.   RU = Russian, TR = Turkish; we calculated the median, because ISEI is an ordinal, not a metric scale.

The frequency of the items was controlled using the Leipzig University's corpus of *Deutscher Wortschatz* (Glück 2011: 32). The items are representative of everyday language.[2] In the expressive test the children had to name an object ("what is this?"), an action ("what is she doing?") or an antonym ("what is the contrary of *hot*"?); in the receptive version they had to point to one of four pictures, showing the item as well as a phonetic distractor, a semantic distractor and a casual distractor. The German normalization of the WWT 6–10 was accomplished in 2004 using a norm sample of 880 monolingual German-speaking children (Glück 2011: 28–30) and a comparative study involving 54 bilingual Turkish-German children (Glück 2011: 19). The testing of the heritage language was carried out by specially trained, bilingual native speakers. Apart from a few exceptions, both languages were tested at an interval of one to four weeks; the language order was chosen randomly. The test took an average of approximately 50 minutes per language and child. In order to achieve optimum data quality, all tests were reviewed using audio recordings of the tests. The children's answers were rated post-test: this means that unexpected but appropriate answers could be accepted. The test was stopped if the children could not provide answers for 5 items in succession (stopping criterion).

A structured interview was conducted with the children. The parents or care persons (henceforth referred to as *parents*) were surveyed using a bilingual questionnaire. The questionnaires allowed information to be collected on, for example, the child's linguistic biography, the languages spoken by the child and the child's parents, literal and cultural activities regarding reading, theater and homework support, and the family's social environment, such as family income, educational level and profession. Concerning the parents' self-assessment of their own language proficiency, the parental questionnaire contained excerpts from descriptions of the CEFR, and the parents were asked in the questionnaire to assess their proficiency in the heritage language. Socio-economic status was surveyed using the parameters of family income, of parents' highest level of education, as well as of learned and practiced profession. The amount of heritage language support varied, in most cases, between 45 and 90 minutes per week, with some rare cases extending to 180 minutes per week.

The impact of variables on heritage language acquisition was analyzed for the Russian and Turkish groups separately. As the dependent variable, we decided not to take the raw values for the vocabulary size in the heritage language but rather the grade-normalized percentile ranks for each group. This solution serves several purposes: it controls for the age differences, the disruption rate and the performance within the communities as well as obtaining a variety from 1 to 100 percentile ranks for the statistical analysis. We paid particular attention to ethical considerations,

---

2.   Controlled for German (frequency classes 7–18) and Turkish (frequency classes 5–19)

informing all the participants and their parents about their rights in Russian, Turkish or German and obtaining informed consent. A moderation analysis was conducted only if the effect of the factors was different in both communities. A mediator analysis was performed if the effects were identical in both communities and the degree of factors differed between the communities. To analyze the moderator impact of the community on the relation between the ISEI and heritage language vocabulary, we used a regression analysis, incorporating the community and the ISEI (for mothers / fathers; for learned profession / current professional activity) as centered main factors, their impact as a moderator factor and the heritage language performance as a dependent variable.

3.2   Findings

3.2.1   *Input patterns and vocabulary*

Input
Regarding the input pattern, the Turkish children are addressed in German far more frequently than the Russian children $\chi^2(5, 198) = 21.99, p = .001$ (Figure 1). 39.8% of the Turkish children are addressed predominantly in German (GER+GER, GER+HL/GER) compared to 19% of the Russian children. 22% of the Turkish children are addressed only in their heritage language compared to 47% of the Russian children.
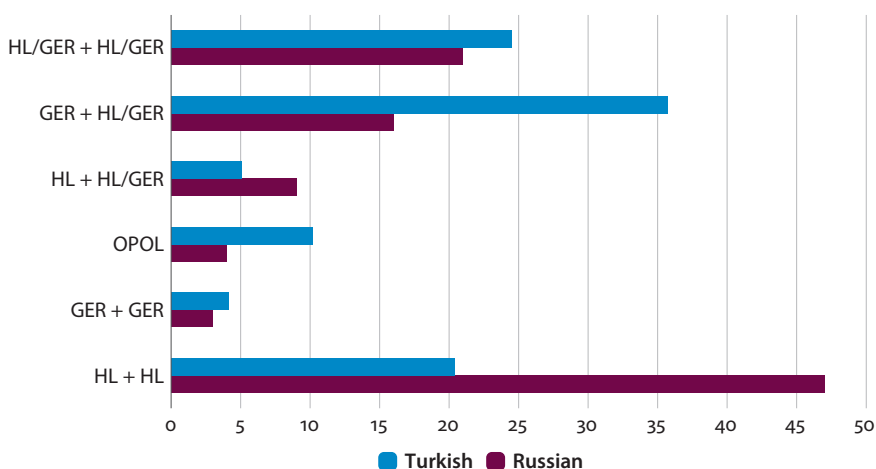
Figure 1. Distribution of input patterns by both parents, ratio of tested children (in %) for the Turkish and Russian communities[3]

---

3.   HL = heritage language, either Russian or Turkish; GER = German; OPOL = one person, one language: one parent uses the heritage language, whereas the other parent uses German with the child.

In our samples, more than 60% of the children receive institutional language support in the heritage language: 62.5% of the Russian-German children and 65.3% of the Turkish-German children enjoy additional heritage language support at school and/or from an association.

3.2.2   *Development of expressive and receptive vocabulary*

Interrupted tests
45.1% of the Russian and 39.8% of the Turkish children failed to complete the expressive test, so the interruption rate is quite high. With the receptive test, the interruption rate in the Russian sample is 2.7% and 10.5% in the Turkish sample.

Expressive vocabulary
Figure 2 shows only the scores for the expressive performance of those children who completed the test. When the expressive performance of the children of the two groups, Turkish and Russian bilinguals, is compared, the t-test for independent samples yields a highly significant finding $t(119) = 17.16$, $p < .001$. Even when the performance is not contrasted across all school grades, but for each grade separately, the t-test yields a significant result each time $ps < .05$. The Russian-speaking children exhibit higher values than the Turkish-speaking children by three standard deviations in the expressive test. In the Turkish group, we find a steady increase from the first to the fourth grade. With $F(3, 55) = 4.91$, the ANOVA yields a highly significant result $p = .004$. In contrast, for the Russian group, a comparison of the average raw values shows a stagnation of the heritage language over the first four school years in the expressive vocabulary. A comparison between grades using an ANOVA yields no significant result $F(3, 57) = 0.35$, $p = .785$.
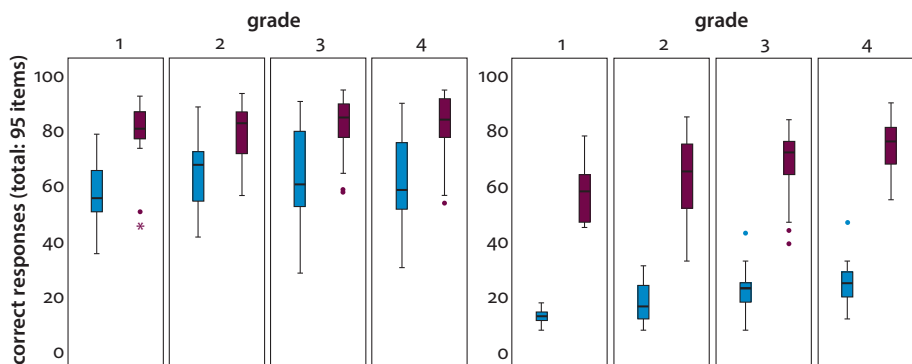
Figure 2. Expressive test results for the Russian (left) and Turkish (right) sample; blue: Expressive, red: Receptive results

Receptive vocabulary

The average results for the receptive vocabulary size are 79.8 (ʀᴜ) and 65.9 (ᴛʀ). This difference is highly significant $t(186) = 7.63$, $p < .001$. The considerable receptive vocabulary size in Russian increases only slightly by four items on average over four years at school; hence the ANOVA shows no significant change in the receptive vocabulary size in Russian $F(3, 97) = .46$, $p = .707$. Similarly, the receptive vocabulary size in Turkish increases slightly by an average of around five items over four years in education, but the ANOVA shows a highly significant change in the receptive vocabulary size in Turkish $F(3, 81) = 4.91$, $p = .003$. We calculated the difference between the average expressive and receptive performance, i.e., words that were recognized but not named; for the Russian children this difference is 21.4 items for the first grade, 15.4 for the second, 19.8 for the third and 20.7 for the fourth. The difference between the average expressive and receptive performance for all children in the Turkish group is much higher: 45.3 items for the first grade, 44.5 for the second, 47.1 for the third and 48.5 for the fourth.

The children who completed the expressive test in Russian achieved a mean value of 85.87 ($SD = 7.17$) in the receptive test, and those who failed to complete the expressive test (hereafter referred to as "dropouts") scored a mean of 72.62 ($SD = 12.28$). Dropouts and children who completed the expressive test do not differ significantly in age. The difference in receptive vocabulary size of children who did the expressive test versus the dropouts is highly significant $t(101) = 6.81$, $p < .001$. If the performance in the receptive Turkish test of children who completed the expressive test is compared to those of the dropouts, a difference in raw values of 16 items across all age groups is revealed, 70.8 vs. 54.8 correct answers. The difference in the receptive performance between dropouts and children who completed the expressive test in Turkish is highly significant $t(83) = 6.23$, $p < .001$, and on average 2.75 items higher than the difference in the Russian group.

Shift

We calculated for each child the difference of the vocabulary in German (= majority language, ML) and the heritage language by subtracting the raw value in the heritage language from the raw value in the ML (Figure 3). Negative values indicate that the heritage language vocabulary is dominant, and positive values indicate that the ML vocabulary is dominant. A shift from the heritage language to a ML dominant vocabulary can be seen in all groups; the ANOVA is significant for Russian $F(3, 57) = 3.72$, $p = .016$ and Turkish $F(3, 55) = 5.77$, $p = .002$.
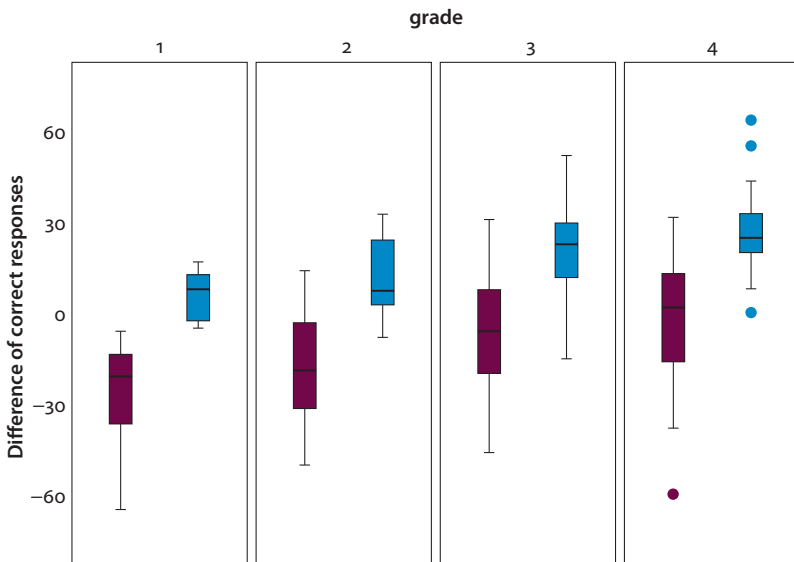
Figure 3. Distribution of difference of raw values for the majority and heritage languages among the four grades of primary school for the Russian (blue) and Turkish (red) sample; includes only children who completed the expressive test in the heritage language

3.2.3   *The impact of input and socio-economic status on vocabulary*

The strongest test results were achieved by those children who lived in an environment where both or at least one parent consistently used the heritage language. Figure 4 shows the impact of input on the expressive test results. Comparing all parental input patterns for Russian speaking children, the ANOVA yields significant results $F(5, 92) = 3.92$, $p = .003$. Russian children who grew up with the input patterns "OPOL" and "HL+HL" (use of the HL by both parents) exhibit the best outcomes in the heritage language. There is also a highly significant effect in the Turkish-speaking sample $F(5, 92) = 5.01$, $p < .001$.

The HL/GER+GER (i.e., one parent speaks both languages, the other parent the ML) pattern shows a marginally significant difference between the performance for Russian and Turkish children $t(49) = -1.84$, $p = .071$.[4] A moderator analysis with the ANOVA was conducted with age as a covariate and performance in the HL, including the children who did not complete the task (performance is zero), as a dependent variable. The analysis revealed, besides the main effects of the community

---

4.   Note that Figure 4 compares percentile ranks in the Russian and the Turkish group, not proficiency values.
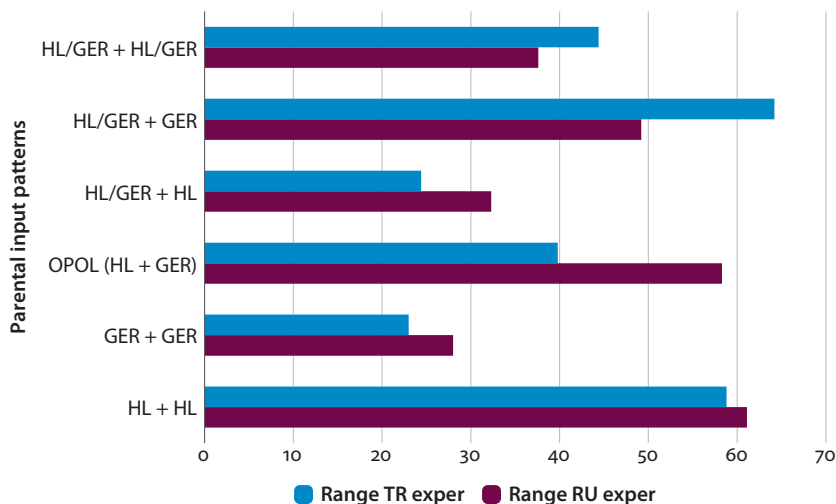
Figure 4. Vocabulary size in percentile ranks in the heritage language in relation to the parental input patterns (blue for Turkish children, red for Russian children)

$F(1, 18) = 13.36$, $p < .001$ and the main effect of the input pattern $F(5, 18) = 4.62$, $p = .001$, a moderator effect of the community $F(5;18) = 2.32$, $p = .045$.

The proficiency of the heritage language of mothers significantly correlates with vocabulary size in the heritage language in children (RU: .299*; TR: .310*). We find a non-significant tendency for fathers' proficiency in the Russian group $r = .199$, whereas in the Turkish group the proficiency in children is independent from the proficiency of fathers $r = .05$, n.s.

Language support from associations or school classes has a positive effect only on the Russian group and not on the Turkish group: Russian bilinguals who receive additional heritage language support from an association or at school have, on average, a significantly larger lexicon of $M = 61.2$ compared to children who do not receive such support: $M = 32.5$ (**), represented in class-normalized percentile ranks. Turkish bilinguals who get heritage language support show, on average, a similar test performance of $M = 51.3$ compared to children without additional support: $M = 53.5$ (n.s.), in class-normalized percentile ranks. The ANOVA analysis reveals, besides the main effects of the community $F(1, 20) = 24.38$, $p < .001$ and institutional support; yes vs. no; $F(1, 20) = 23.52$, $p < .001$, a moderator effect of the community $F(1;20) = 26.29$, $p < .001$.

With respect to socio-economic status, family income has no influence on test performance in the Russian group $F(5, 64) = .86$, $p = .511$ or in the Turkish group $F(5, 44) = .91$, $p = .480$. With respect to the impact of the parents' highest level of education (Figure 5), a university degree has a significant positive effect on

children's vocabulary size of the heritage language Russian: $F(2, 81) = 4.72, p = .012$ for mothers; $F(2, 77) = 5.37, p = .007$ for fathers. The parents' level of education in the Turkish sample has no significant effect on the children's vocabulary size of the heritage language: $F(3, 52) = 1.69, p = .180$ for mothers; $F(3, 48) = .37, p = .774$ for fathers. Again, there is no relation between the attitudes of the father and the vocabulary size of the heritage language in children in the Turkish group. There is also no significant effect for mothers, but we found a slight tendency in the case of mothers, which contrasts to the effect in the Russian group: In the Russian sample, the children of highly qualified mothers are more proficient in their heritage language; in the Turkish sample, this is the case with children whose mothers have no degree (Figure 5). The lack of significance can be ascribed to the small sample of parents without any degree (7.1% of mothers). However, including the community and the highest level of education (for mothers and fathers) as factors in the moderator-analysis with the ANOVA did not reveal any significance for the moderating influence of the community.
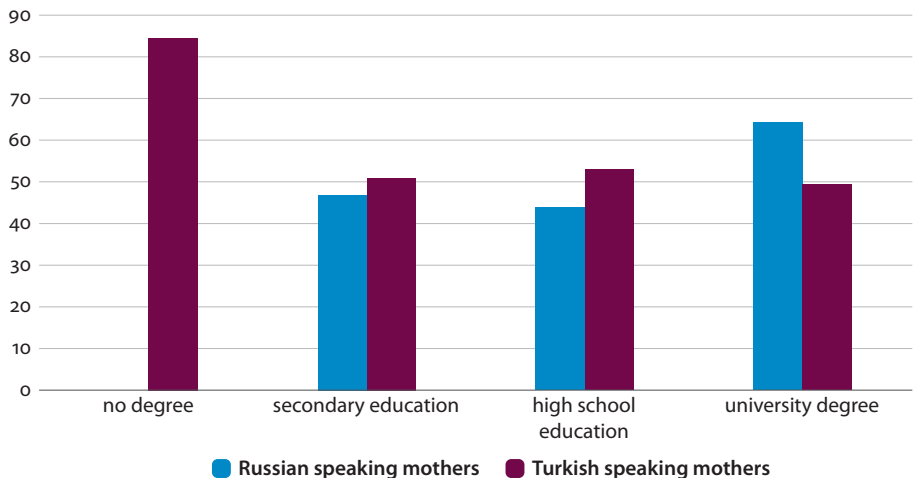
Figure 5. How the mothers' highest qualifications affect their children's test results (vocabulary in grade-normalized percentile rank for group)

For vocabulary acquisition, not only the parents' highest level of education is important, but also where it was obtained. The children's Russian vocabulary scores are higher if one of the two parents obtained their highest qualification in Russia (for mothers: 59.6 vs. 42.3, $t(54) = -2.36, p = .022$; for fathers: 59.2 vs. 45.2, $t(71) = -1.81, p = .074$). The same holds in the case of Turkish skills: these are higher if either the mother or the father completed their highest qualification in Turkey (for mothers: 65.5 vs. 47, $t(78) = -2.59, p = .011$; for fathers: 60.6 vs. 49.2,

$t(50) = 1.41$, $p = .165$). However, the differences with regard to expressive performance are only significant when this applies to the mother.

We revealed a relation between the places in which the highest degree of education was obtained and the heritage language vocabulary of children for both communities (Figure 6). These results point towards a mediator relationship. The mediator analysis with the Sobel-test revealed a partial mediation effect for mothers, $t\_Sobel = 2.98$, $p = .001$, and also for fathers, $t\_Sobel = -2.80$ $p = .005$. So this indirect effect of the community explains, in part, the influence of the community on heritage language acquisition.
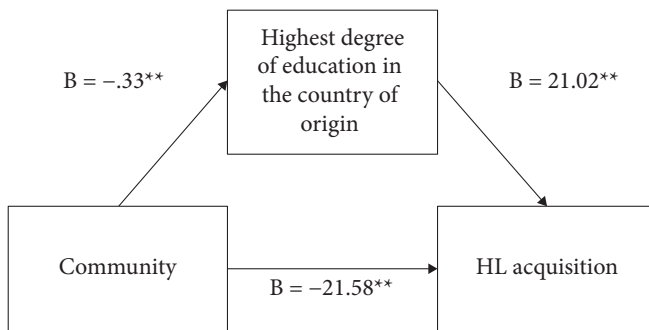
Figure 6. The mediation relationship of the community and the country in which the highest level of education was obtained for mothers and its impact on heritage language acquisition in children

Regarding the impact of the learned and practiced profession parameterized at ISEI, the correlation of Russian vocabulary in grade-normalized percentile ranks and the ISEI for the learned profession in terms of the level of education is highly significant for both the fathers $r = .310**$ and the mothers $r = .307**$. This is also the case for the current professional activity practiced by the fathers $r = .270**$, but not by the mothers $r = .052$ (n.s). In the Turkish group, the correlation of vocabulary in percentile ranks and the ISEI for the learned profession and the current activity in terms of the level of education is neither significant for fathers nor for mothers.

Regarding the learned profession of the mother, we found a main effect for the profession $\beta = .19$, $p = .039$, and for the community $\beta = -.32$, $p = .001$ and a moderator effect $\beta = -.16$, $p = .052$. Regarding the learned profession of the father, we found a main effect for the learned profession $\beta = .22$, $p = .018$, and for the community $\beta = -.32$, $p = .001$ and a moderator effect $\beta = -.22$, $p = .008$.

For the current occupation of the mother, the current professional activity was not significant, the community has a significant impact $\beta = -.41$, $p < .001$, but the moderator effect is not significant. For the current professional activity of the father,

we found a main effect for the current activity $\beta = .17$, $p = .049$, a significant effect for community $\beta = -.42$, $p < .001$ and a moderator effect $\beta = -.16$, $p = .045$. Taken together, we found a moderator effect of the community for the learned profession of the mother and the father as well as for the current activity of the father.

4.   Discussion

Expressive vocabulary in the Russian sample was well developed in the first grade ($M = 57$, $SD = 12.84$), but failed to progress during grades 2–4. We view these findings as an attrition effect, because a constant growth of approximately 8 to 10 words a day in vocabulary can be expected at school age (Klein 2001). The findings contrast with those of Klassert et al. (2014), who identified an increase in vocabulary as the children got older. The difference in the receptive vocabulary size of children who did the expressive test if compared to the dropouts[5] is perhaps not even relevant, considering the big difference in the expressive performance. All of the children understood the majority of the items in Russian. In contrast, the children in the Turkish sample had a considerably poorer expressive command of their heritage language in the first grade ($M = 12$, $SD = 3.7$) and a strong floor effect, but they did exhibit a slight yet continuous increase. The increase of the scores (four versus five items) is similar in both groups, i.e. Turkish and Russian bilinguals, but in the Turkish group the initial values are much lower and consequently the improvement over four years is bigger and also significant. The percentage of children who were unable to perform the expressive test in the Turkish version is high and indicates a lack of mastery of expressive vocabulary. However, all children were highly proficient regarding receptive vocabulary of the heritage language at all grades between 1 and 4. All of the children, even those who failed to complete the expressive part of the test, understood most of the tested words. This means that these word forms and word meanings already existed in their mental lexicons. We also find a shift in the relation between both languages: The older Russian children tended to be balanced bilinguals, whereas the older Turkish speaking children tended to be dominant in the majority language German.

The Turkish and the Russian groups in our sample exhibit differences in their linguistic and social practices, in socio-economic status, in language characteristics (e.g., agglutinative versus inflected language, language contact with German, frequency of items etc.), as well as in their migration history and practices. Both communities differ systematically in some areas, namely the highest achieved

---

5.   Participants who had to quit the test.

qualification, but not in others, such as family income. Linguistic practices, such as the family input patterns, are different, reflecting the values, dynamics and integration courses of migration. Command of the heritage language also differs, particularly in the expressive domain. Community-specific historical and political circumstances lead to different social structures and different linguistic experiences. For the Turkish migration to Europe and Germany, in particular, the *Anwerbeabkommen* of the 1950s, when the German government recruited many workers from Turkey, Italy and other countries, and the following years were highly influential, with the consequence that many parents of bilingual Turkish children today are second or third generation migrants. Thus, they did not have the opportunity to use their first language in education. However, in the Turkish community in Germany it is a frequent practice to marry a partner from Turkey. With respect to the history of contemporary Russian migration, we find many families with two first generation parents who received their education in Russia and used their first language for academic purposes. All these political and social constraints have effects on the kind of language the parents use with their children, on their goals and choices in education and therefore on very important and known factors in language acquisition, such as input, language choice, evaluation of bilingualism and heritage languages as well as language activities at home.

All of this contributes to similarities and differences between the test groups. We find the same effects in both communities in three factors: parental proficiency (mothers: yes, fathers: no), income (no effect), and place where the highest education level was obtained. We cannot confirm a mediator constellation for parental proficiency because both communities do not differ in this respect. We find differences, however, regarding where the parents received their highest education level; thus we find a *mediator* effect of the place of graduation.

We established different effects regarding four factors: input patterns (Russian children have lower heritage language vocabulary scores with mixed input than Turkish children), language support (only positive effect for RU), parents' highest level of education (RU: university degree correlates with best proficiency, TR: no degree), as well as learned and practiced profession ISEI (RU: positive correlation, TR: no correlation). *Moderator* effects have been found regarding family input patterns, the impact of institutional language support, the learned profession of the parents, and the actual occupation of the father. We attribute the different directions of these effects to different beliefs, traditions and shared values in the communities, as supposed by Scheele, Lesemann and Mayo (2010) and Montrul (2016).

Looking at factors, the exposure analysis tells us that it is sufficient and necessary that one person in the family addresses the child continuously in the heritage language for acquisition of the heritage language to be successful (see also Unsworth 2014), and that it is even better if both parents use the heritage language

(De Houwer 2007). Regarding parental input patterns, our results confirm the positive effect of OPOL as a language strategy until school age, offering multiple stimuli in both languages and resulting in a good proficiency in the heritage language (see, e.g., De Houwer et al. 2014). However, the sample sizes for the patterns HL/GER+HL, OPOL and GER+GER are small. Thus, it is difficult to clearly state an impact. The heritage language proficiency of the Russian children decreases with mixed input. In the Turkish community a mixed input has fewer negative effects on heritage language proficiency. We attribute this to intense intergenerational contacts that could contribute to heritage language input.

Additional language support has an impact on some children, but not on all of them. Although parental and institutional input can delay attrition, this does not necessarily lead to a constant and age-adequate growth in expressive vocabulary size (see also Gagarina et al. 2014). One reason why language support does not make a difference in both groups is the heterogeneity of the quality and quantity of additional language support, which sometimes takes place under difficult conditions, for example, as school lessons in the afternoon without a connection to the official curriculum. The quality of the heritage language support should be improved by training teaching staff, ensuring further didactical development, and creating incentives through the use of internationally recognized diplomas. We attribute the relatively high degree of children obtaining heritage language support in our sample to an effect of our recruitment strategy, since we asked schools and associations involved in language support to help us in recruiting the sample.

Regarding the effects of level of education and practiced profession, we found unexpected directions of impact. The children of Turkish mothers with low levels of education exhibit the largest vocabulary in the heritage language. We assume that these mothers use the Turkish language extensively at home and provide rich and consistent heritage language input. However, in the Russian group the children of mothers with a university background show the best command of their HL. Regarding the Russian group, the positive correlation of the education of the mothers with heritage language proficiency of the children confirms the findings of Lambert & Taylor (1996) and Smithson et al. (2014). The differences between the two groups in our sample could be due to different attitudes of the parents towards education and different academic experiences regarding the importance of maintenance of the heritage language versus school success and the necessity to be proficient in the ML.

After having discussed the results, we want to propose some critical reflections on the method we used here. The question remains as to whether the test has the same degree of difficulty in all languages (Mueller-Gathercole 2013), and this could be a limitation of our results. With regard to the analyses carried out here, it becomes clear that the Turkish children on average had more obvious difficulties

in answering the items although the items occur in their environment as often as with the other children. The relatively high number of tested items ($n = 95$) is a quality measure to compensate for possible variation with regard to difficulty of the items for the tested individuals. Furthermore, the children tested here share a social and cultural background, as they all live in Germany. A comparative study which uses the same procedure in France and Turkey deals with this problem extensively (Ertek 2017). This study shows that the test used for Turkish and French monolingual children is more difficult than for German monolingual children. In the third grade, for example, Turkish monolingual children score 55.2, French monolingual children score 52.3 and German monolingual children score 71 (Ertek 2017). However, as far as the heritage language is concerned, the French-Turkish third graders with an average score of 23.6 are stronger than their German-Turkish peers with an average score of 12.7 in testing of expressive vocabulary. It should be noted that the comparison of groups with different multilingual speakers is an exciting and insightful field and provides important information on how different speakers act in the same test setting. At the same time, it cannot be ruled out that the comparative approach suits some constellations better than others.

5.   Conclusion

Due to the risk of misinterpretations arising from comparing groups, our team engaged in numerous critical discussions and took several considerations into account throughout the course of the study. It would be misleading to label some groups as doing better or worse in bilingualism. The discussion surrounding the impact of a community on heritage language maintenance should not be focused on ethnicity, but should take into account the influence of the environment and society on language acquisition. As such, different parents may have very different beliefs and goals when it comes to education and family language policy. This may be partly predetermined or influenced by their social networks, migration history or biography.

There is another critical point to make on the idea of affiliation to a community. In this study, belonging to a community was simply categorized as speaking the same heritage language. This is a helpful construct for statistical analysis, but individuals have weaker or stronger relations to other migrants, they can be active parts of an existing, well-organized community, or they can see themselves as isolated or even so assimilated that there is no visible difference to other people in the society, or they can be anything in between. The notion of affiliation to a community in reality is much more complicated than in this approach.

This leads to the conclusion that networks in migration situations should also be included in the analysis of the multilingual context. Understanding communities as networks with political, historical, biographical and linguistic accords, as well as looking at networks and communities is an important issue for a more differentiated understanding of multilingualism in migration.

Funding information

References

Aichinger, G. (ed.). 1997. *Migration und soziale Netzwerke: Studien in Favelas in Rio de Janeiro und Sao Bernardo do Campo*. Pfaffenweiler: Centaurus.

Andrews, D. R. 1999. *Sociocultural Perspectives on Language Change in Diaspora. Soviet Immigrants in the United States* [IMPACT: Studies in Language and Society 5]. Amsterdam: John Benjamins. https://doi.org/10.1075/impact.5

Bialystok, E., Luk, G., Peets, K. F. & Yang, S. 2010. Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition* 13: 525–531. https://doi.org/10.1017/S1366728909990423

Brehmer, B. & Usanova, I. 2015. Let's fix it? Contact-induced changes in word order patterns of Russian heritage speakers in Germany. In *Transfer Effects in Multilingual Language Development* [Hamburg Studies on Linguistic Diversity, 4], H. Peukert (ed), 159–186. Amsterdam: John Benjamins.

Brizić, K. 2007. *Das geheime Leben der Sprachen: Gesprochene und verschwiegene Sprachen und ihr Einfluss auf den Spracherwerb in der Migration*. Münster: Waxmann.

De Houwer, A. 2007. Parental language input patterns and children's bilingual use. *Applied Psycholinguistics* 28: 411–424. https://doi.org/10.1017/S0142716407070221

De Houwer, A., Bornstein, M. H., Putnick, D. L. 2014. A bilingual-monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics* 35: 1189–1211. https://doi.org/10.1017/S0142716412000744

Ertek, B. 2017. Développement du vocabulaire en turc et en français. PhD dissertation, Université du Rouen, Rouen.

Gagarina, N. et al. 2014. Age, input quantity and their effect on linguistic performance in the home and societal language among Russian-German and Russian-Hebrew preschool children. In *The Challenges of Diaspora Migration: Interdisciplinary Perspectives on Israel and Germany*, R. Silbereisen, P. Titzmann & Y. Shavit (eds), 63–82. Farnham: Ashgate.

Ganzeboom, H. B. G. & Treiman, D. J. 2003. Three internationally standardised measures for comparative research on occupational status. In *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables*, J. H. P. Hoffmeyer-Zlotnik & C. Wolf (eds), 159–193. Dordrecht: Kluwer. https://doi.org/10.1007/978-1-4419-9186-7_9

Glück, C. W. 2011. *Wortschatz- und Wortfindungstest für 6- bis 10-Jährige*, 2nd edn. Munich: Urban & Fischer.

Gogolin, I., Dirim, İ., Klinger, T., Lange, I., Lengyel, D., Michel, U., Neumann, U., Reich, H. H., Roth, H.-J. & Schwippert, K. 2011. *Förderung von Kindern und Jugendlichen mit Migrationshintergrund FÖRMIG. Bilanz und Perspektiven eines Modellprogramms* Münster: Waxmann.

Gurak, D. T. & Caces, F. 1992. Migration networks and the shaping of migration systems. In *International Migration Systems: A Global Approach*, M. M. Kritz, L. L. Lim & H. Zlotnik (eds), 150–176. Oxford: Clarendon Press.

Haug, S. 2000. *Klassische und neuere Theorien der Migration*. Mannheim: Zentrum für Europäische Sozialforschung.

Hoff, E., Core, C., Place, S., Rumiche, R., Senor, M. & Parra, M. 2012. Dual language exposure and early bilingual development. *Journal of Child Language* 39: 1–27. https://doi.org/10.1017/S0305000910000759

Hoff, E. & Tian, C. 2005. Socioeconomic status and cultural influences on language. *Journal of Communication Disorders* 38: 271–278. https://doi.org/10.1016/j.jcomdis.2005.02.003

Isurin, L. & Ivanova-Sullivan, T. 2008. Lost in between: The case of Russian speakers. *Heritage Language Journal* 6: 72–104.

Klassert, A., Kauschke, C. & Gagarina, N. 2014. Object and action naming in Russian- and German-speaking monolingual and bilingual children. *Bilingualism: Language and Cognition* 17(1): 73–88. https://doi.org/10.1017/S1366728913000096

Klein, W. 2001. Typen und Konzepte des Spracherwerbs. In *Deutsch als Fremdsprache*, G. Helbig, L. Götze, G. Henrici & H. Krumm (eds), 604–616. Berlin: De Gruyter.

Kondo-Brown, K. 2005. Differences in language skills: Heritage language learner subgroups and foreign language learners. *The Modern Language Journal* 89: 563–581. https://doi.org/10.1111/j.1540-4781.2005.00330.x

Kupisch, T., Belikova, A., Özçelik, Ö., Stangen, I. & White, L. 2017. Restrictions on definiteness in the grammars of German-Turkish heritage speakers. *Linguistic Approaches on Bilingualism* 7(1): 1–32. https://doi.org/10.1075/lab.13031.kup

Lambert, W. E. & Taylor, D. M. 1996. Language in the lives of ethnic minorities: Cuban american families in Miami. *Applied Linguistics* 17: 477–500. https://doi.org/10.1093/applin/17.4.477

Lee, J. 2011. Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics* 32: 69–92. https://doi.org/10.1017/S0142716410000299

Milroy, L. 1980. *Language and Social Networks*. 1st edn. New York NY: Blackwell.

Montanari, E. 2017. *Beschulung von neu in das niedersächsische Bildungssystem zugewanderten Schülerinnen und Schülern in der Sekundarstufe I*. Hildesheim: Stiftung Universität Hildesheim.

Montrul, S. 2008. *Incomplete Acquisition in Bilingualism. Re-examining the Age Factor* [Studies in Bilingualism 39]. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.39

Montrul, S. 2016. *The Acquisition of Heritage Languages*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139030502

Mueller-Gathercole, V. C. 2013. *Issues in the Assessment of Bilinguals*. Bristol: Multilingual Matters.

Pearson, B., Fernández, S. C. & Oller, K. D. 1993. Lexical development in bilingual infants and toddlers: Comparison to momolingual norms. *Adjunct Faculty Author Gallery* 7: 93–120.

Pfaff, C. W. 1994. Early bilingual development of Turkish children in Berlin. In *The Cross-Linguistic Study of Bilingual Development*, G. Extra & L. Verhoeven (eds), 75–97. Amsterdam: Koninklijke Nederlandse Akademie van Wetenschappen.

Putnam, M. T. & Sánchez, L. 2013. What's so incomplete about incomplete acquisition? A pro-legomenon to modeling heritage language grammars. *Linguistic Approaches to Bilingualism* 3: 478–508. https://doi.org/10.1075/lab.3.4.04put

Rehbein, J., Herkenrath, A. & Karakoç, B. 2009. Turkish in Germany – on contact-induced language change of an immigrant language in the multilingual landscape of Europe. *STUF* 62: 171–204.

Scheele, A. F., Lesemann, P. P. M. & Mayo, A. Y. 2010. The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics* 31: 117–140. https://doi.org/10.1017/S0142716409990191

Schweizer, T. 1996. *Muster sozialer Ordnung*. Berlin: Reimer.

Shin Kim, M. 2013. The mental lexicon of low-proficiency Korean heritage learners. *Heritage Language Journal* 10: 17–35.

Siemund, P., Gogolin, I., Schulz, M. E. & Davydova, J. (eds). 2013. *Multilingualism and Language Diversity in Urban Areas: Acquisition, Identities, Space, Education* [Hamburg Studies on Linguistic Diversity 1]. Amsterdam: John Benjamins. https://doi.org/10.1075/hsld.1

Smithson, L., Paradis, J. & Nicoladis, E. 2014. Bilingualism and receptive vocabulary achievement: Could sociocultural context make a difference? *Bilingualism: Language and Cognition* 17(4): 810–821. https://doi.org/10.1017/S1366728913000813

Sürig, I., Şimşek, Y., Schroeder, C. & Boneß, A. 2016. *Literacy Acquisition in School in the Context of Migration and Multilingualism: A Binational Survey* [Hamburg Studies on Linguistic Diversity, 5] Amsterdam: John Benjamins. https://doi.org/10.1075/hsld.5

UNCRC. 1989. *United Nations Convention of the Rights of a Child*. <https://www.ohchr.org/en/professionalinterest/pages/crc.aspx> (9 December 2018).

Unsworth, S. 2014. Amount of exposure as a proxy for dominance in bilingual language acquisition. In *Language Dominance in Bilinguals: Issues of Measurement and Ope-rationalization*, C. Silva-Corvalán & J. Treffers-Daller (eds), 156–173. Cambridge: CUP.

Vetter, E. 1997. *Nicht mehr Bretonisch? Sprachkonflikt in der ländlichen Bretagne*. Frankfurt: Peter Lang.

Heritage and non-heritage bilinguals

The role of biliteracy and bilingual education

Maria Andreou[i], Ifigenia Dosi[ii], Despina Papadopoulou[ii]
and Ianthi Maria Tsimpli[iii]
[i]University of Cologne / [ii]Aristotle University of Thessaloniki /
[iii]University of Cambridge

The present study explores the effects of literacy support in the languages of the bilingual child on the linguistic and cognitive skills of (non-)heritage speakers. 70 children speaking Albanian (L1) and Greek (L2) are divided into three groups according to whether they receive literacy support in their L1, L2 or in both languages. To test the children's proficiency in Greek, we used an expressive vocabulary task, two working memory tasks, a non-verbal intelligence task and a sentence repetition task (SRT). The findings suggest that good levels of biliteracy established through bilingual education positively influence the child's linguistic and cognitive performance. Furthermore, the lack of working memory effects on the children's SRT performance emphasizes the importance of biliteracy development and its educational support, which together with vocabulary knowledge contribute to bilingual development.

Keywords: (non-)heritage speakers, biliteracy, educational setting, working memory, sentence repetition task

1. Introduction

Heritage speakers form a rather heterogeneous bilingual group, since they may differ with respect to language profiles and histories as well as levels of literacy. Quite often the majority language, i.e. the language of the society and schooling, overtakes the minority language, i.e. the home language, and becomes dominant. Moreover, the minority language, which is the heritage speakers' first language (L1), rarely receives further support during the school years. However, a number of studies (Andreou 2015; Dosi, Papadopoulou and Tsimpli 2016) have highlighted the significance of L1 support in education, since it has been observed that when

both languages are equally supported and developed, cognitive advantages are more pronounced. Apart from biliteracy the role of educational setting (submersion bilingualism, immersion bilingualism, heritage bilingual education and balanced education) has also been pointed out in recent studies (Andreou & Tsimpli in press, Dosi & Papadopoulou 2019). Hence, bilingual education seems to boost linguistic (Andreou & Tsimpli in press, Dosi, Papadopoulou & Tsimpli 2016) and cognitive abilities (Andreou 2015, Dosi 2016).

Turning to the tools employed to measure language abilities, a commonly used tool is the Sentence Repetition Task (SRT). However, it is not yet clear whether performance on SRT is affected by working memory abilities. Some researchers argue that memory skills have a significant impact on SRT performance (Alloway & Gathercole 2005), whereas other researchers do not find any correlation between SRTs and working memory abilities (Klem, Melby-Lervaog, Hagtvet, Lyster, Gustafsson & Hulme 2015).

The present study explores Albanian-Greek heritage and non-heritage bilingual speakers' language skills in each language using SR tasks and seeks to investigate the contribution of working memory to accuracy scores. Furthermore, it is investigated whether (a) L1 support and (b) the development of the two languages within – or outside – the school framework affects the bilinguals' linguistic behavior.

2.   Theoretical background

2.1   Heritage speakers

What makes heritage speakers unique is that the language of first socialization is not developed in the same way as for other bilinguals and that they become dominant in the majority language (Montrul 2010). Heritage language bilinguals are initially exposed to the minority language. From school age, however, the majority language, which is also the language of education, takes over and in adolescence and adulthood, the majority language becomes dominant (Montrul 2009). Usually the proficiency level of the home language, i.e. the minority language, is rather low, while proficiency in the majority language is high.

One factor that has been found to have an impact on minority language attrition is Age of Onset (AoO) in bilingualism (Montrul 2008a). According to Montrul (2002) incomplete acquisition of the L1 is usually greater in simultaneous bilinguals and in early sequential bilinguals than in late sequential bilinguals, even though these findings are not supported by some studies (Kaltsa, Tsimpli & Rothman 2015; Tsimpli et al. 2004). Hence, lexical retrieval (for a review see Ecke 2004), but also inflectional morphology (for Spanish: Montrul 2008a,b; for Russian: Polinsky

2007, 2008) in the L1 are challenging for heritage speakers. For example, gender performance has been found to be problematic in the minority language of both Spanish-American and Russian-American heritage speakers (for Spanish: Montrul 2008b; for Russian: Polinsky 2008). Additionally, tense and aspect in the minority language have been shown to be vulnerable as well (for Russian: Polinsky 2007; for Spanish: Silva-Corvalán 1994). Another vulnerable characteristic in the heritage speakers' minority language is the domain of the interfaces and more specifically pronominal resolution (Tsimpli & Sorace 2006). Core syntax, by contrast, seems to be rather intact (Iverson 2012).

Another important factor that affects heritage speakers' language performance is whether they have received support in their L1. Thus, when heritage speakers have not received L1 support, they are more likely to face difficulties in their L1 and even underperform in several aspects of language compared to L2 learners (Rothman 2007).

2.2   Biliteracy

Biliteracy refers to the bilinguals' ability to read and write in two languages. However, to date very few studies have investigated the effects of biliteracy in the development of grammar and vocabulary. A small body of studies has revealed that biliterate bilinguals outperform monoliterate bilinguals and monolinguals in phonological awareness tasks and reading fluency measures (Schwartz, Leikin & Share 2005; Leikin, Schwartz & Share 2010). This is unsurprising because these tasks have been shown to be good predictors of reading abilities in monolingual and bilingual children (Ehri et al. 2001). However, it is unclear whether biliteracy also has a positive impact on vocabulary and grammatical abilities. More recent studies have pointed out that not only biliteracy, but also the educational setting plays a significant role in bilinguals' cognitive and language performance (Andreou 2015; Dosi 2016; Dosi & Papadopoulou 2019). Thus, educational settings that support both languages boost the advantages of bilingualism in the linguistic and cognitive domain. Similarly, in a study by Dosi et al. (2016) it has been observed that biliterate bilinguals from a bilingual educational setting exhibited higher performance with respect to their cognitive updating abilities compared to monoliterate bilinguals from a monolingual educational setting. Interestingly, language skills, i.e. performance on verbal aspect, were also positively affected by the biliterates' cognitive performance, since no differences were observed in the linguistic task between the two bilingual groups, although the biliterate participants lagged behind the monoliterate ones in vocabulary knowledge.

2.3   Sentence repetition task

The Sentence Repetition Task (SRT) is considered as an appropriate tool for measuring general language abilities. Many studies (Ellis 2005; Erlam 2006) have shown that "SRTs tap into the learner's implicit knowledge" (Marinis & Armon-Lotem 2015). Despite the fact that SRTs have been used as an assessment and a diagnostic tool cross-linguistically for clinical populations, such as children with SLI, only recently have SRTs been used with bilingual children (Chondrogianni, Andreou, Nerantzini, Varlokosta & Tsimpli 2013; Kaltsa, Prentza & Tsimpli 2019). Studies using SRTs with bilingual children have shown that AoO and Length of Exposure (LoE) are reliable predictors of the bilinguals' performance (Chiat, Armon-Lotem, Marinis, Polisenska, Roy & Seeff-Gabriel 2013).

According to Klem et al. (2015), SRT is best seen as a complex linguistic task that reflects the integrity of language processing systems at many different levels, such as speech perception, lexical (vocabulary) knowledge, grammatical skills and speech production. Studies on L2 have pointed out that, apart from processing, SRTs tap into language representations and provide an indication of general language proficiency (Chaudron & Russell 1990; Munnich, Flynn & Martohardjono 1994).

As pointed out in the Introduction, there is a long discussion about the involvement of working memory abilities in SRTs. Some studies have indicated that the contribution of memory (either working memory or short-term memory) on accurate performance is crucial (Alloway & Gathercole 2005; Alloway, Gathercole, Willis & Adams 2004). Furthermore, very few studies have claimed that the contribution of memory is more important than the contribution of language to SRT performance (Hamayan, Saegert & Larudee 1977). By contrast, other studies have not observed any involvement of memory in SRT performance (Okura & Lonsdale 2012; Dosi et al. 2016). Nonetheless, there is a great number of recent studies which maintain that SRTs draw on both language ability and cognitive resources, primarily working memory (Riches 2012; Klem et al. 2015) and especially in relatively short sentences (Fattal, Friedmann & Fattal-Valevski 2011), since language processing is less demanding and memory abilities are more prevalent in those sentences (Alloway et al. 2004).

2.4   Similarities and differences between Greek and Albanian[1]

Greek and Albanian are Indo-European as well as south Balkan languages and share many linguistic properties.[2] Both Greek and Albanian are morphologically rich languages as the nominal and verbal paradigms are inflected for a number of morphosyntactic features. The nominal paradigm of the two languages is marked for number, case and gender, as shown in examples (1) and (2). However, the determiner in Greek comes before the noun (example (1)), whereas in Albanian the determiner follows the noun (example (2)).

(1)   o   ántras
       the.NOM.MASC.SG man.NOM.MASC.SG

(2)   hoteli
       hotel.the.NOM.MASC.SG

Moreover, the two systems present differences in terms of case marking. In Greek there are four cases (nominative, genitive, accusative, vocative), while in Albanian there are five cases (nominative, genitive, accusative, dative, ablative).

In terms of gender the two systems show some differences. The Greek nominal system has a three-way gender distinction (masculine, feminine and neuter nouns), whereas the neuter gender in Albanian is limited.

Regarding number, both languages distinguish between singular and plural, as indicated in examples (3) (Greek) and (4) (Albanian).

(3)   a.   o   ántras
             the.NOM.MASC.SG man.NOM.MASC.SG
       b.   oi   ántres
             the.NOM.MASC.PL   man.NOM.MASC.PL

(4)   a.   studenti
             student.the. NOM.MASC.SG
       b.   studentët
             student.the. NOM.MASC.PL

---

1.   The Albanian examples (5), (7), (15) are from Revithiadou & Spyropoulos (2013), while examples (9) to (12) are from Dosi (2016: 37).

2.   This section does not intend to present an exhaustive linguistic comparison between the two languages (see Revithiadou & Spyropoulos 2013). Rather, it aims at presenting some key features of the two languages which are also present in the experimental task employed.

Additionally, definite and indefinite articles are inflected for gender, in Albanian, while adjectives are inflected for gender and number but not case. Only determiners are marked for case (see example (5)). On the contrary in Greek both determiners and adjectives are marked for gender, number and case (see example (6)).

(5) a. djali   të   mirë
         child.the.DAT.MASC.SG a.DAT.MASC.SG nice.MASC.SG
     b. djalë   të   mire
         child.the.ACC.MASC.SG a.ACC.MASC.SG nice.MASC.SG

(6) a. énas   kalós   ánthropos
         a.NOM.MASC.SG nice.NOM.MASC.SG man.NOM.MASC.SG
     b. énan   kaló   ánthropo
         a.ACC.MASC.SG nice.ACC.MASC.SG man.ACC.MASC.SG

Furthermore, differences are detected in the order of the adjectives relative to the noun: In Albanian the adjective comes after the noun, whereas in Greek the adjective typically precedes the noun (see examples (7) and (8)).

(7) Erdhën   djemtë   e   mirë.
     come.PAST.ACT.IND.3PL boy.NOM.MASC.PL the nice.MASC.PL
     "The good boys came."

(8) Írthan   oi   kaloí
     come.PAST.ACT.IND.3PL the.NOM.MASC.PL good.NOM.MASC.PL
     mathités.
     student.NOM.MASC.PL
     "The good students came."

Turning to the verbal system, both languages mark verbs for voice, mood, tense, aspect, person and number (for Greek see examples (9) and (11); for Albanian see examples (10) and (12)). Regarding verbal aspect, Greek and Albanian distinguish between perfective (see examples (9) and (10) respectively) and imperfective (see examples (11) and (12) respectively), while Albanian further marks progressivity by means of the prefix *po* (see example (12)), though this seems to be dialect dependent (Bertinetto, Ebert & de Groot 2000).

(9) Zográfise   énan   pínaka.   (Greek)
     Paint.PERF.PAST.ACT.IND.3SG a.ACC.MASC.SG picture.ACC.MASC.SG
     "(S)he painted a picture."

(10) Dje   vajza   lexo**i**
yesterday girl.the.NOM.FEM.SG read.PERF.PAST.ACT.IND.3SG
librin.   (Albanian)
book.the.ACC.MASC.SG
"Yesterday the girl read the book."

(11) Xtes   óli méra   zográfize   énan
yesterday the whole day painted.IMP.PAST.ACT.IND.3SG a.ACC.MASC.SG
pínaka.   (Greek)
picture.ACC.MASC.SG
"Yesterday (s)he was painting a picture the whole day."

(12) Dje   vajza   (po) lexo**nte**
yesterday girl.the.NOM.FEM.SG read.IMP.PAST.ACT.IND.3SG
librin   gjithë   mëngjes.
book.the.ACC.MASC.SG whole.MASC.SG morning.the.ACC.MASC.SG
"Yesterday the girl was reading the book the whole morning."   (Albanian)

Another similarity between the two languages is that both have clitic pronouns. In Greek clitic doubling (CD) and clitic left dislocation (CLLD) are possible, as in (13a) and (13b) respectively. Drachman (1983) reports that clitic doubling is also allowed in the Albanian language (see example (14)).

(13) a. Ton   kafé   ton
      the.ACC.MASC.SG coffee.ACC.MASC.SG him.ACC.MASC.SG
      ípie   viastiká o
      drink.PAST.ACT.IND.3SG quickly the.NOM.MASC.SG
      papoús   sto   kafeneío.   [CLLD] (Greek)
      grandpa.NOM.MASC.SG in the coffee shop.ACC.NEUT.SG

    b. O   papoús   ton
      the.NOM.MASC.SG grandpa.NOM.MASC.SG him.ACC.MASC.SG
      ípie   viastiká ton   kafé
      drink.PAST.ACT.IND.3SG quickly the.ACC.MASC.SG coffee. ACC.MASC.SG
      sto   kafeneío.   [CD] (Greek)
      in the coffee shop.ACC.NEUT.SG
      "The grandfather drank his coffee in the coffee shop in a hurry."

(14) Kopshtari   e   mbolli
gardener.the.NOM.MASC.SG it.ACC.MASC.SG plant.PAST.PERF.ACT.IND.3SG
kopshtin   e   xhaxhait   tim me
garden.the.ACC.MASC.SG it.ACC.MASC.SG uncle.ACC.MASC.SG my with
qershia   të   vogla.   (CD) (Albanian)
cherry.ACC.FEM.PL the.ACC.FEM.PL small.FEM.PL
"The gardener planted my uncle's garden with small cherry-trees."

The two languages differ with regard to infinitives, since infinitives are present in Albanian (*për të larë*, i.e. 'to wash'), whereas infinitives do not exist in Greek (see Revithiadou & Spyropoulos 2013). Greek has complement clauses headed by a number of complementizers, such as *na, óti, pu* etc., and Albanian also has complement clauses headed by *te, çka, që* etc. (Newmark, Hubbard & Prifti 1982). Moreover, the two languages have a number of conjunctions introducing adverbial clauses expressing a number of relations, such as time, cause, purpose etc. In both languages there are also relative clauses; the relative pronoun agrees with the head noun it refers to in terms of gender, number and case (see (15), Newmark, Hubbard & Prifti 1982).

(15) a. Ne po vizitojmë muzeun, në
We. NOM.1PL will visit.FUT.ACT.IND.1PL museum.the.ACC.MASC.SG in
të cilin ka shumë
the which.ACC.MASC.SG there.are. PRES.ACT.IND.3PL many
piktura të bukura.
picture.NOM.FEM.PL the beautiful.FEM.PL

b. tha episkeftoúme to mouseío
visit. FUT.ACT.IND.1PL the.ACC.NEUT.SG museum.ACC.NEUT.SG
sto opoío ipárxoun
in.the.which.ACC.NEUT.SG there are.PRES.ACT.IND.3PL
poloí ómorfoi pínakes.
many.NOM.MASC.PL beautiful.NOM.MASC.PL picture.NOM.MASC.PL
"We will visit the museum, in which there are many beautiful paintings."

Thus, Greek and Albanian, notwithstanding their differences, share many structural features, such as rich inflectional nominal and verbal systems, clitics and clitic doubling, as well as subordination markers.

3. The present study

The present study aims to explore (a) the predictors of SRT for (non) heritage bilingual children, and (b) the role of literacy in SRT performance in (non) heritage bilinguals, i.e. whether the heritage bilinguals who receive (written) language support in their minority language (i.e. Albanian) differ from the heritage bilinguals who do not receive support in their L1 and, additionally, whether the two heritage groups differ from the Albanian-Greek bilinguals who receive language support in both languages.

3.1   Participants

The participants of the present study are children speaking Albanian (L1) and Greek (L2) and are divided into three groups: (a) Albanian-Greek heritage speakers whose L1 literacy is not supported (henceforth; HS1 n = 31), (b) Albanian-Greek heritage speakers whose L1 literacy is supported (henceforth; HS2 n = 10) and, finally, (c) Albanian-Greek bilinguals[3] who receive support for literacy in both languages (henceforth; BL n = 29). For the heritage groups the country of residence is Greece, whereas for the bilingual group the country of residence is Albania. The participants of HS groups are members of the Albanian immigrant community in Greece. On the contrary, the BL group's parents are Albanian who immigrated to Greece, bore and raised their children in Greece and afterwards went back to Albania due to the financial crisis in Greece. Nevertheless they stay in contact with Greece since they have relatives and seasonal jobs in Greece.

All participants were between 8 and 12 years old (for HS1: mean 10.6, SD 1.1; for HS2: mean 10.5, SD 1.1; for BL: mean 10.6, SD 1.3). At this point, it should be clarified how biliteracy is determined in this study. The HS1 group attended Greek state schools, was exclusively taught through the Greek language and did not receive literacy in Albanian. The HS2 group attended a supplementary school (in Greece) where they attended Albanian courses (for 2 hours/week) and, thus, their literacy in Albanian is enhanced. The BL group attended a bilingual educational setting, where both languages were supported: Albanian was the main medium of instruction, 23 hours/week, while Greek was taught for 14–16 hours/week.

Further information about the participants was obtained through a child questionnaire (Mattheoudakis, Chatzidaki & Maligkoudi 2014). The main questions were grouped in three categories: (a) home language history, (b) current language use and (c) early literacy practices. *Home language history* refers to exposure to each language from birth up to the age of schooling (i.e. up to the age of six). *Current language use* refers to the language preferences for daily activities (i.e. memorizing phone numbers, calculating, telling the time or watching TV), oral interaction with family members and friends and the language that they feel they understand or speak better. *Early literacy practices* refer to activities such as shared-book reading

---

3.   An anonymous reviewer raised the issue that the HS2 and the BL groups receive literacy in both of their languages; thus, the difference between them is the country of residence and the intensity of literacy instruction in Albanian. Indeed, heritage speakers are bilinguals with varying degrees of exposure to their two languages and (bi)literacy (see also Rothman 2007 and Montrul 2013 for a similar point). In this paper, however, following previous studies (see Introduction), we used the term heritage speakers (HS) for those Greek-Albanian bilinguals who learnt their heritage (minority) language within their family while they reside in Greece and are exposed and educated in the majority language (Greek).

in preschool age. The results from the questionnaires reveal that, with respect to *home language history*, no differences were detected among the three groups ($p > .1$). Regarding *early literacy practices* in Greek no differences were attested either ($p > .1$). On the other hand, the three groups were found to differ with respect to *early literacy practices* in Albanian [$F(2, 66) = 4.510$, $p = .015$]. Post hoc tests (Bonferroni) have shown that the BL group was more frequently exposed to shared-book reading practices in Albanian than the HS1 group ($p = .025$), whereas no other differences were attested. With regard to *current language use* differences were observed in the use of both Greek and Albanian [$F(2, 67) = 16.477$, $p < .001$; $F(2, 67) = 18.104$, $p < .001$, respectively]. Bonferroni's post hoc tests have revealed that the BL group uses Greek less frequently in their everyday life compared to the two heritage groups ($p < .001$ and $p = .001$, for HS1 and HS2, respectively); in contrast, they use Albanian more often compared to the two heritage groups ($p < .001$, for both comparisons). Within group comparisons have shown that all groups differ with respect to current language use. Both heritage groups use Greek more often than Albanian; however, the differences between the use of each language were more prevalent in the heritage group with no L1 support [for HS1: $t(30) = 2.901$, $p = .007$; for HS2: $t(9) = 2.277$, $p = .049$]. In the BL group the pattern is different too, since they use Albanian more frequently than Greek [$t(26) = -4.927$, $p < .001$].

Finally, the socioeconomic status of the children was measured by maternal education (in accordance with Ensminger & Fothergill 2003; Hoff, Laursen & Tardif 2002) and it was calculated on a 5-point Likert-type scale, with 5 representing the highest educational level attained from compulsory primary education to tertiary education, which we adapted from the UBILEC (see also Unsworth 2013). Our results indicated that there was a significant group effect [$F(2, 67) = 16.877$, $p < .001$]; the BL group's mothers had significantly more years of education than the HS groups' mothers (HS1: $p < .001$; HS2 group: $p < .001$). By contrast, no significant differences ($p > .950$) were detected between the HS1 and HS2 children (see Table 1 for more details on the groups' ages and SES).

Table 1. Participants' biodata (means and SDs)

Group	Age	SES
	M (SD)	M (SD)
HS1	10.6	3.3
($n = 31$)	(1.1)	(1.37)
HS2	10.5	2.6
($n = 10$)	(1.1)	(0.77)
BL	10.6	4.4
($n = 29$)	(1.3)	(0.35)

3.2   Materials and procedure

All three groups undertook two sentence repetition tasks, one in each language. Additionally, we employed independent measures of expressive vocabulary, as a means of testing the children's language proficiency, which were also used as predictors of the children's performance in the sentence repetition tasks. Finally, all participants were also tested for their cognitive abilities: non-verbal intelligence and (non) verbal working memory. The verbal working memory task was in Greek. In addition, the instructions of the cognitive tasks were given only in Greek.

3.2.1   *Expressive vocabulary tests in Greek and Albanian*

The Greek (Vogindroukas, Protopapas & Sideridis 2009; adaptation from Renfrew 1997) and the Albanian (Kapia & Kananaj 2013) vocabulary tests assess expressive vocabulary in Greek and Albanian. Both tests are naming tasks and consist of 50 items each, arranged in order of increasing difficulty.

In both tests the same procedure has been followed; the examiner presents a picture of an object which the child is expected to name. Testing stops when the child either completes all trials or provides wrong naming (or no response) in five consecutive trials. The highest possible score for each test is 50, with each correct naming response given 1 point.

3.2.2   *Non-verbal intelligence (fluid intelligence)*

To measure children's nonverbal abilities, we used the Raven's Coloured Progressive Matrices (Raven, Court & Raven 2008). Given that this test has not been standardized for Greek-speaking children, we used the raw scores for further statistical comparisons since the children were matched on age. The children were asked to complete three levels of twelve test items each, consisting of visuo-spatial conceptual matching exercises, which were increasing in difficulty. Each correct answer was scored as 1 point and there were no penalty points for wrong identifications. The maximum score was 36 points.

3.2.3   *Verbal working memory task*

The backwards digit recall task is a computerized measure of verbal working memory from the Automated Working Memory Assessment (AWMA, Alloway 2007), which was normed for Greek by Chrysohoou (2006). In this test the child is required to recall a sequence of spoken digits in reverse order. Digit sequences were audiotaped by a native speaker of Greek with 1-second distance between the offset of a digit and the onset of the next one. Test reliability of the AWMA is reported in Alloway (2007) and test validity in Alloway, Gathercole, Kirkwood & Elliott (2009). This is a span task in which the number of digits to remember increases progressively over successive blocks containing 6 trials each. The criterion for moving on to the

next block was the correct recall of 4 out of the 6 trials. Testing stopped if the child failed in 3 trials in one block. The task consisted of 6 blocks, starting with 2-digit sequences in the first block and increasing to 7-digit sequences in the last block.

Procedure

Each child was tested individually in a quiet room. In order to familiarize the child with the task, three practice sessions were administered in which the child had to recall 2- and 3-digit sequences. During practice sessions, children were given feedback for both accurate and inaccurate digit recalls. The examiner entered the scores in an evaluation grid.

Scoring

Each correct answer was given 1 point and there were no penalty points for wrong recalls. The first 4 consecutive successful recalls in each block were scored as 6 points; the child then moved to the next block. If the fourth correct recall was on the fifth or the sixth trial, the child got in total 5 and 4 points, respectively. The same scoring procedure was repeated across all 6 blocks. The highest possible score was 36 points.

3.2.4   *Non-verbal working memory task*

Non-verbal working memory was assessed through the Rotating Figure task (Alloway 2007). This is a very demanding task, since the participant must simultaneously process and temporarily store visuo-spatial information. The participant is shown a picture of two Mr. X figures and has to identify whether Mr. X with the blue hat is holding the ball in the same hand as Mr. X with the yellow hat. Mr. X with the blue hat may also be rotated. At the end of each trial the child has to recall the location of each ball in Mr. X's hand in sequence, by pointing to a picture with eight compass points. The procedure is similar to the one of the previous task.

Scoring

Each correct answer is given 1 point while there are no penalty points for wrong answers. For each level (span) there are 6 trials, which equals to 6 points for the corresponding number of correct answers. The first 4 consecutive successful trials in each level award 6 points and the right to move on to the second level. If the fourth correct answer is trial 5, the child gets in total 5 points and moves to the second level, if it is trial 6 s/he gets a total of 4 points and moves on to the next level. The same procedure is repeated in all 7 levels. The discontinuation rule applies when the child gives 3 wrong answers in any of the 7 levels and the procedure is terminated, but in this study we use the total score on the task and not the level the participant reached. The highest score for correct trial responses is 42, and for span is 7. The span calculation is based on the 4 correct responses in the last level reached by each child.

3.2.5   *Sentence repetition task*

The SRT employed in this study was designed by COST Action IS0804 (Chondrogianni et al. 2013) and adapted to each language. In both languages all sentences were grammatical and no fillers were included. Additionally, all sentences were taped and auditorily presented. The Greek SRT consisted of 32 sentences, containing a variety of structures such as negation, clitics, coordination, complement and relative clauses, wh-questions and adverbials.[4] The Albanian SRT comprised 60 sentences, which involved structures similar to those in the Greek SRT.

Procedure

During the task the child listened to each sentence only once and repeated it as accurately as possible. There was a practice session, so that the participants became familiar with the procedure. The participants listened to the sentences via headphones and their responses were recorded.

Coding

The task was assessed with respect to two factors, (a) grammaticality and (b) accuracy. The grammaticality scores referred to whether the utterance of the participant was grammatical or not. Thus, if the utterance produced by the participant was grammatical, (s)he received 1 point, while, if the utterance was ungrammatical, the participant received no points (see examples (16a and b)).

(16) a. Ton   kafé   ton
          the.ACC.MASC.SG coffee.ACC.MASC.SG him.ACC.MASC.SG
          ípie   viastiká o
          drink.PAST.ACT.IND.3SG quickly the.NOM.MASC.SG
          papoús   xtes   sto   kafeneío.
          grandpa.NOM.MASC.SG yesterday in the coffee shop.ACC.NEUT.SG
          "The grandfather drank the coffee hastily yesterday at the coffee shop."
          (Grammatical: 1 point)

      b. *O   kafés   ton
          the.NOM.MASC.SG coffee.NOM.MASC.SG him.ACC.MASC.SG
          ípie   viastiká o
          drink.PAST.ACT.IND.3SG quickly the.NOM.MASC.SG
          papoús   xtes   sto   kafeneío.
          grandpa.NOM.MASC.SG yesterday in the coffee shop.ACC.NEUT.SG
          "The grandfather drank the coffee hastily yesterday at the coffee shop."
          (Ungrammatical: 0 point)

---

4. See Kaltsa, Prentza & Tsimpli (2019) for the use of the same task with monolingual and Albanian-Greek bilingual children in Greece focusing on structure-related performance.

By contrast, accuracy scores pertained to how accurately the participant repeated the sentence. If the participant's utterance exactly matched the sentence given, the participant received 3 points, whereas, if the participant made any lexical or grammatical substitution, omission or addition, they received 2 points. Moreover, if the participant made two of the aforementioned errors, they received 1 point and, if the participant made three or more errors, they received no point (see examples (17a, b and c)).

(17)   a.   Ton   kafé   ton
             the.ACC.MASC.SG coffee.ACC.MASC.SG him.ACC.MASC.SG
             ípie   viastiká o
             drink.PAST.ACT.IND.3SG quickly the.NOM.MASC.SG
             papoús   xtes   sto kafeneío.
             grandpa.NOM.MASC.SG yesterday In   the coffee shop.ACC.NEUT.SG
             "The grandfather drank the coffee hastily yesterday at the coffee shop."
                                                                         (3 points)
        b.   Ton   kafé   ípie   viastiká
             the.ACC.MASC.SG coffee.ACC.MASC.SG drink.PAST.ACT.IND.3SG quickly
             o   papoús   xtes   sto
             the.NOM.MASC.SG grandpa.NOM.MASC.SG yesterday in the
             kafeneío.   (omission of the clitic *ton*)
             coffee shop.ACC.NEUT.SG   (2 points)
        c.   Ton   kafé   ípie
             the.ACC.MASC.SG coffee.ACC.MASC.SG drink.PAST.ACT.IND.3SG
             o   papoús   xtes   sto
             the.NOM.MASC.SG grandpa.NOM.MASC.SG yesterday in the
             kafeneío.   (omission of the clitic *ton*)
             coffee shop.ACC.NEUT.SG   (omission of the adverb *viastiká*)

Furthermore, the sentences were also coded in detail for any lexical or grammatical types of errors (i.e. errors in function words, in verbal/nominal inflection and word order errors); however, this detailed error analysis will not be presented in the present study.

The SRTs in the two languages differ with respect to the total scores on grammaticality and accuracy, as they contain a different number of sentences. In the Greek SRT the total grammaticality score was 32, since the Greek task has 32 sentences; whereas in the Albanian SRT the total grammaticality score was 60, since the Albanian version has 60 sentences. Regarding the total score on accuracy, this was 96 (32 × 3) in the Greek task, while it was 180 (60 × 3) in the Albanian SRT. In order for the two tasks to be comparable we transformed all scores into percentages.

4. Results

In order to detect possible differences among the groups we ran one-way ANOVA analyses. As dependent variables the scores on each task were set whereas the three groups (BL, HS1, HS2) formed the independent variable. Whenever a significant group effect was detected, we ran post hoc Bonferroni tests to determine which groups were significantly different from each other.

4.1 Expressive vocabulary test in Greek

A one-way ANOVA with group as the between-group variable revealed a significant group effect [$F(2, 67) = 7.230, p = .001$]. Post hoc tests (Bonferroni) showed that the BL group scored significantly lower (29.3) than HS1 (36.5) and HS2 (38.0) children ($p = .004$ and $p = .017$, respectively). There was no significant difference between HS1 and HS2 groups (see Table 2).

4.2 Expressive vocabulary test in Albanian

The results on this task also revealed a significant group effect [$F(2, 67) = 6.863$, $p = .002$]. Post hoc tests (Bonferroni) showed that the BL group scored significantly higher (37.3) than HS1 (31.5) and HS2 (29.4) groups ($p = .006$ and $p = .024$, respectively) (see Table 2).

Table 2. Participants' performance on vocabulary tasks (mean and SD)

Group	Vocabulary in Greek	Vocabulary in Albanian
	M (SD)	M (SD)
HS1	36.5	31.6
($n = 31$)	(7.6)	(7.5)
HS2	38.0	29.4
($n = 10$)	(6.3)	(8.1)
BL	29.4	37.3
($n = 29$)	(9.3)	(5.8)

4.3 Non-verbal intelligence (fluid intelligence)

A group effect was detected in this data set as well [$F(2, 67) = 4.478$, $p = .015$], in that the BL children scored significantly higher (27.1) than the HS1 (23.5) group ($p = .014$), as revealed by the post hoc tests (Bonferroni), whereas no other differences are detected (see Table 3).

4.4 Verbal working memory task

A significant group effect [$F(2, 67) = 10.490$, $p < .001$] was also observed, due to which the BL group scored significantly higher (16.2) than the HS1 (11.0) group ($p < .001$). No other differences were detected between the other groups (see Table 3).

4.5 Non-verbal working memory task

The results of this task were parallel to the ones from the *non-verbal intelligence* task; a group effect was found ($F(2, 67) = 4.982$, $p = .010$), because once again the BL children scored significantly higher (15.0) than the HS1 (11.4) group ($p = .010$). No other differences were detected between the other groups (see Table 3).

Table 3. Participants' performance on cognitive tasks (mean and SD)

Group	Non-verbal intelligence	Verbal working memory	Non-verbal working memory
	M (SD)	M (SD)	M (SD)
HS1	23.5	11.0	11.4
(*n* = 31)	(4.8)	(3.0)	(3.4)
HS2	26.3	13.0	15.0
(*n* = 10)	(4.8)	(3.5)	(5.7)
BL	27.1	16.2	15.7
(*n* = 29)	(4.4)	(5.6)	(6.7)

4.6 Sentence repetition task

In order to detect possible differences among the groups we again conducted one-way ANOVA analyses, following the same procedure as for the previous tasks. The scores on grammaticality and accuracy for each language were set as dependent variables in this task.

Regarding the grammaticality scores, differences were detected among groups in the Greek task ($F(2, 67) = 6.024$, $p = .004$). Post hoc tests (Bonferroni) showed that interestingly the BL groups outperformed the HS1 group ($p = .003$). No other differences were observed. In the Albanian task no differences were detected among the groups ($F(2, 67) = 2.729$, $p = .073$).

Turning to the accuracy scores, a significant group effect was also exhibited ($F(2, 67) = 10.377$, $p < .001$) in the Greek task. More specifically, the Bonferroni test revealed that the BL children were significantly more accurate than the two HS groups ($p = .001$ and $p = .002$, for the HS1 and the HS2 groups respectively). No differences between the two heritage groups were observed. A group effect was attested in the accuracy scores in the Albanian test ($F(2, 67) = 23.537$, $p < .001$). Similarly to what has been observed above, the BL group manifested higher performance compared to the two other groups (HS1: $p < .001$; and HS2: $p = .002$), whereas no other differences were manifested.
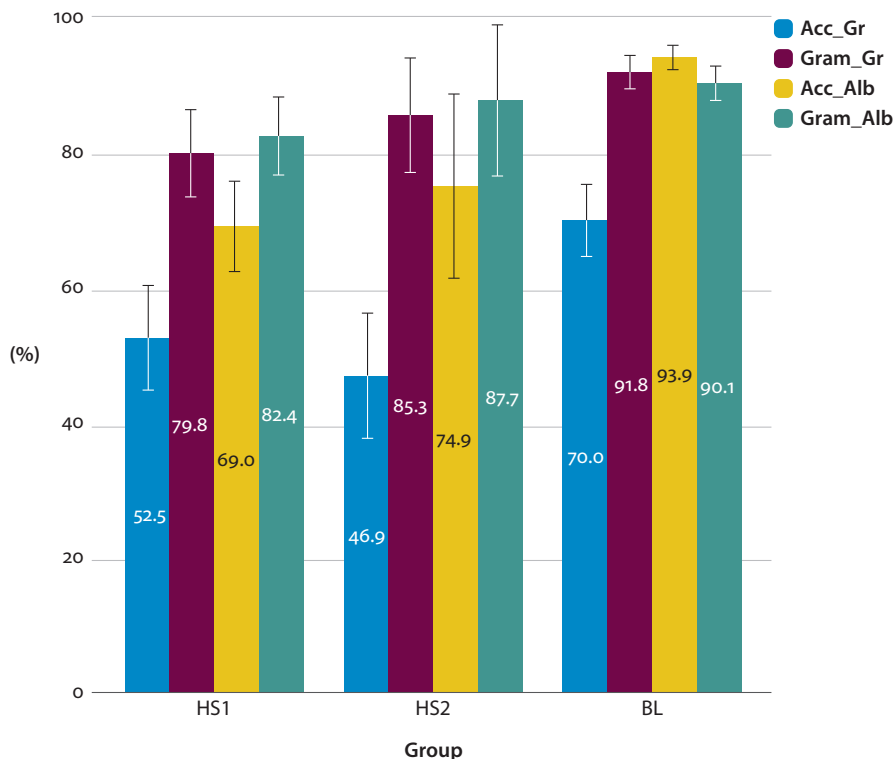
Figure 1. Participants' grammaticality and accuracy scores in the SRTs

We further compared the grammaticality and accuracy scores between the two languages, by performing paired samples t-tests for each group. Interestingly, in terms of grammaticality all groups performed similarly in the Greek and in the Albanian task (for HS1: $t(30) = −.735$, $p = .468$; for HS2: $t(9) = −.387$, $p = .708$; for BL: $t(28) = 1.056$, $p = .300$). In terms of accuracy scores all groups produced the sentences in the Albanian SRT more accurately compared to the Greek SRT (for HS1: $t(30) = −3.867$, $p = .001$; for HS2: $t(9) = −4.118$, $p = .003$; for BL: $t(28) = −10.931$, $p < .001$).

Trying to determine which factors explain the participants' performance on grammaticality scores, we ran linear stepwise regression analyses, one for the Greek version and one for the Albanian version. In the first analysis, grammaticality scores on the Greek SRT were set as dependent variable and chronological age of the participants was set as the only independent variable, since it was the only variable that correlated with the grammaticality scores in this task ($r = .322$, $p = .007$). The results indicated that the age of the participants was predicting 12.5% of participants' performance [$R^2 = .125$, $F(1, 59) = 24.003$, $p < .005$; $\beta = .353$]. By contrast, in the Albanian data set of the grammaticality scores, age_1 and vocabulary scores (in Albanian)$_2$ were the independent variables, since these were the only variables that correlated with the grammaticality scores in the Albanian task ($r_1 = .246$, $p_1 = .043$; $r_2 = .374$, $p_2 = .002$). The results revealed that the vocabulary scores in Albanian were the only predictor of the grammaticality scores in the Albanian SRT [$R^2 = .311$, $F(1, 60) = 27.045$, $p < .001$; $\beta = .557$]. With respect to the accuracy scores in the Greek SRT, educational setting$_1$ and vocabulary scores in Greek$_2$ were set as independent variables ($r_1 = .443$, $p_1 < .001$; $r_2 = .375$, $p_2 = .004$). The results demonstrate that both variables predict the accuracy scores in the Greek SRT [$R^2 = .304$, $F(1, 67) = 7.368$, $p = .008$; $\beta_1 = .604$, $p_1 < .001$ and $\beta_2 = .305$, $p_2 = .008$, respectively]. In the analysis of the accuracy scores from the Albanian SRT, we set as independent variables age_1, scores on non-verbal fluid intelligence$_2$, vocabulary scores (in Albanian)$_3$, early literacy practices in Albanian$_4$, current language use in Albanian$_5$ and educational setting$_6$, since all these variables were found to correlate with accuracy scores in the Albanian task ($r_1 = .437$, $p_1 = .001$; $r_2 = .283$, $p_2 = .028$; $r_3 = .600$, $p_3 < .001$; $r_4 = .356$, $p_4 = .007$; $r_5 = .400$, $p_5 = .002$; and $r_6 = .687$, $p_6 < .001$). The results show that educational setting and vocabulary scores predict the accuracy scores in the Albanian SRT [$R^2 = .570$, $F(1, 58) = 19.831$, $p < .001$; $\beta_6 = −.507$, $p_6 < .001$ and $\beta_3 = .409$, $p_3 < .001$, respectively].

4.6.1   *Syntactic errors (Greek sentence repetition task)*

To examine whether children differed in terms of the syntactic errors they produced, we ran ANOVAs. There were no significant differences between the groups, suggesting that the children produced equal proportions of syntactic errors.

The results revealed that the CLLD/CD structures (44.9%) were more problematic than Subject-Verb-Object (SVO), coordinations (COORD), adverbial clauses (ADVCL), wh-questions (WHQ) sentences ($F(1, 7) = 18.001$, $p = .001$). In particular, post hoc tests using the Bonferroni correction attested the following differences: CLLD/CD vs. SVO ($p = .001$); CLLD/CD vs. COORD ($p = .031$); CLLD/CD vs. ADVCL ($p = .03$); and CLLD/CD vs. WHQ ($p = .06$). For instance, in example (13a), some of the participants omitted the clitic *ton*. Additionally, NEG sentences (see example (18); the participants produced 18b rather than 18a) and COMPCL (see example (19); they produced 19b instead of 19a) were also difficult for the participants ($p > .2$ and $p > .4$, respectively).

(18) a.   O   zográfos   thélei   na
          the.NOM.MASC.SG painter.NOM.MASC.SG want.PRES.ACT.IND.3SG to
          min piánoun   oi   fíloi
          not touch.PRES.ACT.CONJ.3PL the.NOM.MASC.PL friend.NOM.MASC.PL
          tou   tous   pínakés   tou.
          his.GEN.MASC.SG the.ACC.MASC.PL painting.ACC.MASC.PL his GEN.MASC.SG
          "The painter wants his friends not to touch his paintings."

      b.   O   zográfos   den thélei
          the.NOM.MASC.SG painter.NOM.MASC.SG not want.PRES.ACT.IND.3SG
          na piánun   oi   fíloi
          to touch.PRES.ACT.CONJ.3PL the.NOM.MASC.PL friend.NOM.MASC.PL
          tou   tous   pínakés
          his. GEN.MASC.SG the.ACC.MASC.PL painting.ACC.MASC.PL
          tou.
          his. GEN.MASC.SG
          "The painter does not want his friends to touch his paintings."

(19) a.   I   giagiá
          the.NOM.FEM.SG grandmother.NOM.FEM.SG
          thimótan   **óti** se aftá   ta
          remember.PAST.IMPF.ACT.IND.3SG that in these the.ACC.NEUT.PL
          méri   petoúsan   períerga
          place.ACC.NEUT.PL fly.PAST.IMPF.ACT.IND.3PL unusual.NOM.NEUT.PL
          pouliá.
          bird.NOM.NEUT.PL
          "The grandmother remembered that unusual birds were flying around these parts."

b. I   giagiá
    the.NOM.FEM.SG grandmother.NOM.FEM.SG
    thimótan   **pos** se aftá   ta
    remember.PAST.IMPF.ACT.IND.3SG that in these the.ACC.NEUT.PL
    méri   petoúsan   períerga
    place.ACC.NEUT.PL fly. PAST.IMPF.ACT.IND.3PL unusual.NOM.NEUT.PL
    pouliá.
    bird.NOM.NEUT.PL
    "The grandmother remembered that unusual birds were flying around
    these parts."

For all groups, there were more substitutions than omissions and more omissions
than additions (*p* < .001 in both cases).

5.   Discussion

Before proceeding to the discussion of our data, we will provide a short summary
of our findings. Our data revealed that the HS groups outperformed the BL group
in the Greek vocabulary test, while the opposite finding has been attested in the
Albanian vocabulary test, which is expected since the HS groups live in Greece and
their input in Albanian is limited. Furthermore, the BL group's non-verbal intelli-
gence and working memory were significantly higher than those of the HS, while
the BL scored significantly better than the HS group that did receive L1 support
in the verbal working measure. Regarding the SRT scores, the BL group displayed
significantly better accuracy than the HS in both the Greek and the Albanian tasks.
The BL group's grammaticality scores, by contrast, did not significantly differ from
those of the two HS groups in the Albanian task. In the Greek task, though, the BL
group manifested higher grammaticality scores than the HS group which did not
receive L1 support. Additionally, as far as the grammaticality scores are concerned,
no differences were detected between performance on the Greek and the Albanian
test in either group. On the other hand, all groups' accuracy on the Albanian task
was significantly better than that in the Greek task. Finally, age predicted the groups'
performance on grammaticality in the Greek SRT, whereas vocabulary was the
predicting factor of the participants' grammaticality scores in the Albanian task. In
contrast, both vocabulary and educational setting explained the children's accuracy
in both the Greek and the Albanian tasks.

First of all, it is worth noticing that the BL group manifested better cognitive
abilities than the HS groups, which is in line with previous studies (Leikin, Schwartz
& Share 2010). Also, the HS group that did not receive any L1 support displayed
the lowest scores in the verbal working memory task. These findings suggest that

the development of biliteracy and its support by means of education are significant for the improvement of the bilingual child's cognitive abilities. Our results further indicate that biliteracy, and not bilingualism itself, enhances the development of cognitive skills (Dosi et al. 2016).

Another interesting finding is that the HS group that did not receive any L1 support exhibited lower grammaticality scores than the BL group in the Greek task, even though the HS1 group's vocabulary knowledge in Greek exceeded that of the BL group. This result also shows that the lack of L1 support and its further development negatively affects the children's performance even in the L2. As suggested by Cummins (2001), language education in more than one language confers advantages, since several skills can be transferred across the languages that an individual speaks. Similarly, Baker (1993) argues that children developing literacy in two languages are expected to have better oral language abilities. Thus, we maintain that there is a blocking of transfer to majority language skills, due to the absence of the written support of HS1 children's L1 language.

The three groups, however, do not show any differences regarding their grammaticality scores in the Albanian task. In other words, the lack of – or limited – L1 support did not have a significant impact on the production of grammatical sentences in the minority language. We think that this finding is a reflection of the children's pure grammatical knowledge in Albanian, which, after all, is their first language. Notice, though, that the HS groups' accuracy in the Greek and the Albanian SRTs are significantly worse than that of the BL group. This result indicates that, in monolingual educational settings, the bilinguals' ability to produce grammatical sentences in the L1 remains rather intact, whereas their ability to accurately reproduce sentences in their two languages deteriorates. Hence, language abilities more related to literacy, such as the incorporation of lexical, morphosyntactic and semantic indices that is required for accurate sentence repetition, are more likely to manifest a decline, when biliteracy is not reinforced by the educational system.

Turning to the parameters that predicted children's performance on the SRT, the first noteworthy point is that vocabulary knowledge has a significant impact on the bilingual children's SRT performance. Similar results on the role of vocabulary on performance in the same SRT are reported in Kaltsa, Prentza & Tsimpli's (2019) study although the link between vocabulary and both accuracy and grammaticality scores is only found for simultaneous bilingual Albanian-Greek children. SRT accuracy scores in sequential bilinguals do not show a link with vocabulary scores whereas grammaticality scores do. The importance of lexical development on the bilinguals' linguistic performance has been pointed out by a number of recent studies (Andreou 2015; Dosi 2016; Dosi et al. 2016). Such findings imply the interrelatedness between vocabulary and grammatical abilities in bilingual development, an issue requiring further investigation so that it is clarified which grammatical

aspects are more affected by lexical development. The only exception in our data, in which vocabulary did not predict the children's SRT performance, was their grammaticality scores in Greek. Age was the parameter that better explained the bilinguals' performance in this case. So, developmental effects are in play as far as grammaticality in the L2 is concerned; this is not a surprising result keeping in mind that higher age also means longer exposure to the L2. In addition, the educational setting is another important contributor to the bilinguals' accuracy in both languages. Recent findings demonstrate that good levels of biliteracy, established through bilingual education, positively influence the child's linguistic and cognitive performance in a number of tasks, such as quality of connected oral and written discourse (i.e. higher use of causal and temporal connectives, see Andreou 2015), better linguistic (i.e. performance on aspect: Dosi et al. 2016) and cognitive abilities as well as updating skills (Andreou 2015; Dosi 2016). These findings have important consequences for the educational policy, as they prioritize bilingual over monolingual educational programs, which promote the development of the oral and written skills in both languages (Tsimpli 2017).

A final point that needs to be discussed concerns the fact that the children's (verbal and non-verbal) working memory abilities did not predict their performance on the SRT. This result confirms findings of other studies (Klem et al. 2015; Dosi et al. 2016) in which working memory skills did not affect SRT performance. As Marinis and Armon-Lotem (2015) have recently suggested, such findings clearly indicate that performance on SRT is not just an issue of good memory skills, but rather an issue of implicit linguistic knowledge. Thus, if a speaker has not fully acquired a structure, s/he is not able to produce it by means of sheer memorization. We think that the lack of working memory effects on the children's SRT performance further emphasizes the importance of biliteracy development and its educational support, which, together with vocabulary knowledge, contribute to bilingual development.

Acknowledgements

References

Alloway, T. P. 2007. What can phonological and semantic information tell us about the mechanisms of immediate sentence recall? *Memory* 15: 605–615.
https://doi.org/10.1080/09658210701450877

Alloway, T. P. & Gathercole, S. E. 2005. Working memory and short-term sentence recall in young children. *European Journal of Cognitive Psychology* 17: 207–220.
https://doi.org/10.1080/09541440440000005

Alloway, T. P., Gathercole, S. E., Kirkwood, H. J. & Elliott, J. E. 2009. Evaluating the validity of the Automated Working Memory Assessment. *Educational Psychology* 7: 725–734.

Alloway, T. P., Gathercole, S. E., Willis, C. & Adams, A. M. 2004. A structural analysis of working memory and related cognitive skills in early childhood. *Journal of Experimental Child Psychology* 87: 85–106. https://doi.org/10.1016/j.jecp.2003.10.002

Andreou, M. 2015. The Effects of Bilingualism on Verbal and Non-verbal Cognition: The Micro- and Macro-structure of Narratives in the Weak and the Dominant Language of the Bilingual Child. PhD dissertation, Aristotle University of Thessaloniki.

Andreou, M. & Tsimpli, I. M. in press. Bilingualism, biliteracy and syntactic complexity: the role of crosslinguistic influence and cognitive skills. *Language Acquisition, Processing and Bilingualism, TRT7.*

Baker, C. 1993. *Foundations of Bilingual Education and Bilingualism*. Clevedon: Multilingual Matters.

Bertinetto, P. M., Ebert, K. & de Groot, C. 2000. The progressive in Europe. In *Tense and Aspect in the Languages of Europe*, Ö. Dahl (ed.), 517–558. Berlin: Mouton de Gruyter.

Chaudron, C. & Russell, G. 1990. The status of elicited imitation as a measure of second language competence. Paper presented at the *Ninth World Congress of Applied Linguistics*, Thessaloniki, Greece.

Chiat, S., Armon-Lotem, S., Marinis, T., Polisenska, K., Roy, P. & Seeff-Gabriel, B. 2013. The potential of sentence imitation tasks for assessment of language abilities in sequential bilingual children. In *Bilinguals and Assessment: State of the Art Guide to Issues and Solutions from around the World*, V. C. Mueller-Gathercole (ed.), 56–89. Bristol: Multilingual Matters.
https://doi.org/10.21832/9781783090105-005

Chondrogianni, V., Andreou, M., Nerantzini, M., Varlokosta, S. & Tsimpli, I. M. 2013. *The Greek Sentence Repetition Task*. COST Action IS0804.

Chrysochoou, E. 2006. Working Memory Contributions to Children's Listening Comprehension in the Preschool and Elementary School Years. PhD dissertation, Aristotle University of Thessaloniki.

Cummins, J. 2001. *Language, Power and Pedagogy: Bilingual Children in the Crossfire*. Clevedon: Multilingual Matters.

Dosi, I. 2016. The Verbal Aspect in Bilingual Children: The Effect of Linguistic, Cognitive and Environmental Factors [in Greek]. PhD dissertation, Aristotle University of Thessaloniki.

Dosi, I. & Papadopoulou, D. (2019). The role of educational setting in the development of verbal aspect and executive functions: evidence from Greek-German bilingual children, *International Journal of Bilingual Education and Bilingualism*.
https://doi.org/10.1080/13670050.2018.1539465

Dosi, I., Papadopoulou, D. & Tsimpli, I. M. 2016. Linguistic and cognitive factors in elicited imitation tasks: A study with mono- and biliterate Albanian-Greek bilingual children. In *Proceedings of the 40th Annual Boston University Conference on Language Development*, J. Scott & D. Waughtal (eds), 101–115. Somerville MA: Cascadilla Press.

Drachman, G. 1983. *Parameters and Functions of Clitic Doubling in Modern Greek*. Ms, Aristotle University of Thessaloniki.

Ecke, P. 2004. Language attrition and theories of forgetting: A cross-disciplinary review. *International Journal of Bilingualism* 8: 321–354.
https://doi.org/10.1177/13670069040080030901

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z. & Shanahan, T. 2001. Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly* 71(3): 393–447.

Ellis, N. C. 2005. At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition* 27: 305–352.
https://doi.org/10.1017/S027226310505014X

Ensminger, M. E. & Fothergill, K. E. 2003. A decade of measuring SES: what it tells us and where to go from here. In *Socioeconomic Status, Parenting and Child Development*, M. H. Bornstein & R. H. Bradley (eds), 13–27. Mahwah NJ: Lawrence Erlbaum Associates.

Erlam, R. M. 2006. Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics* 27(3): 464–491. https://doi.org/10.1093/applin/aml001

Fattal, I., Friedmann, N. & Fattal-Valevski, A. 2011. The crucial role of thiamine in the development of syntax and lexical retrieval: A study of infantile thiamine deficiency. *Brain* 134(6): 1720–1739. https://doi.org/10.1093/brain/awr068

Hamayan, E., Saegert, J. & Larudee, P. 1977. Elicited imitation in second language learners. *Language and Speech* 20(1): 86–97. https://doi.org/10.1177/002383097702000109

Hoff, E., Laursen, B. & Tardif, T. 2002. Socioeconomic status and parenting. In *Handbook of Parenting, Vol. 2: Biology and Ecology of Parenting*, 2nd edn, M. Bornstein (ed), 231–252. Mahwah NJ: Lawrence Erlbaum Associates.

Iverson, M. 2012. On L1 Attrition and the Interface Hypothesis-A Case Study of Spanish Attrition in Contact with Brazilian Portuguese. PhD dissertation, University of Iowa.

Kaltsa, M., Prentza, A. & Tsimpli, I. M. 2019. Input and literacy effects in simultaneous and sequential bilinguals: The performance of Albanian-Greek speaking children in sentence repetition. *International Journal of Bilingualism*. https://doi.org/10.1177/1367006918819867

Kaltsa, M., Tsimpli, I. M. & Rothman, J. 2015. Exploring the source of differences and similarities in L1 attrition and heritage speaker competence: Evidence from pronominal resolution. *Lingua* 164: 266–288. https://doi.org/10.1016/j.lingua.2015.06.002

Kapia, E. & Kananaj, A. 2013. *The Albanian Vocabulary Instrument*. Tiranë: Center for Albanian Studies, Institute of Linguistics and Literature.

Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S. A. H., Gustafsson, J. E. & Hulme, C. 2015. Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science* 18(1): 146–154. https://doi.org/10.1111/desc.12202

Leikin, M., Schwartz, M. & Share, D. L. 2010. General and specific benefits of bi-literate bilingualism: a Russian-Hebrew study of beginning literacy. *Reading and Writing* 23: 269–292. https://doi.org/10.1007/s11145-009-9210-x

Marinis, T. & Armon-Lotem, S. 2015. Sentence repetition. In *Methods for Assessing Multilingual children: Disentangling Bilingualism from Language Impairment*, S. Armon-Lotem, J. de Jong & N. Meir (eds), 95–122. Bristol: Multilingual Matters.

Mattheoudakis, M., Chatzidaki, A. & Maligkoudi, C. 2014. Bilingual education and types of bilingualism. Paper presented at the Workshop Language Knowledge and Development in Native and Non-native Speakers. Thessaloniki: Aristotle University of Thessaloniki.

Montrul, S. 2002. Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition* 5(1): 39–68. https://doi.org/10.1017/S1366728902000135

Montrul, S. 2008a. *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor* [Studies in Bilingualism 39]. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.39

Montrul, S. 2008b. Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning* 58: 503–553. https://doi.org/10.1111/j.1467-9922.2008.00449.x

Montrul, S. 2009. Incomplete acquisition of tense-aspect and mood in Spanish heritage speakers. *The International Journal of Bilingualism* 13(3): 239–269. https://doi.org/10.1177/1367006909339816

Montrul, S. 2010. Current issues in heritage language acquisition. *Annual Review of Applied Linguistics* 30: 3–23. https://doi.org/10.1017/S0267190510000103

Montrul, S. 2013. Bilingualism and the heritage language speaker. In *The Handbook of Bilingualism and Multilingualism*, 2nd edn, T. K. Bhatia & W. C. Ritchie (eds), 168–189. Malden MA: Wiley-Blackwell.

Munnich, E., Flynn, S. & Martohardjono, G. 1994. Elicited imitation and grammaticality judgment tasks: What they measure and how they relate to each other. In *Research Methodology in Second Language Acquisition*, E. Tarone, S. Gass & A. Cohen (eds), 227–245. Mahwah NJ: Lawrence Erlbaum Associates.

Newmark, L., Hubbard, P. & Prifti, P. 1982. *Standard Albanian: A Reference Grammar for Students*. Stanford CA: Stanford University Press.

Okura, E. & Lonsdale, D. 2012. Working memory's meager involvement in sentence repetition tests. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, N. Miyake, D. Peebles & R. P. Cooper (eds), 2132–2137. Austin TX: Cognitive Science Society.

Polinsky, M. 2007. Incomplete acquisition: American Russian. *Journal of Slavic Linguistics* 14: 191–262.

Polinsky, M. 2008. Russian gender under incomplete acquisition. *The Heritage Language Journal* 6: 40–71.

Raven, J., Court, J. & Raven, J. C. 2008. *Raven's Coloured Progressive Matrices and Vocabulary Scales*. London: Pearson Education.

Renfrew, C. E. 1997. *Action Picture Test*, 4th edn. Bicester: Winslow.

Revithiadou, A. & Spyropoulos, V. 2013. *A comparative study of Albanian-Greek grammatical structures. Project 'Education of Immigrant and Repatriate Students', Operational Program of Education and Lifelong Learning (NSRF 2007–2013)*. Aristotle University of Thessaloniki Publications, Research Committee, Thessaloniki. <http://www.diapolis.auth.gr/index.php/2013-10-17-09-02-51/-52> (5 October 2018).

Riches, N. G. 2012. Sentence repetition in children with specific language impairment: An investigation of underlying mechanisms. *International Journal of Language & Communication Disorders* 47(5): 499–510. https://doi.org/10.1111/j.1460-6984.2012.00158.x

Rothman, J. 2007. Heritage speaker competence differences, language change and input type: Inflected infinitives in heritage Brazilian Portuguese. *International Journal of Bilingualism* 11(4): 359–389. https://doi.org/10.1177/1367006907011004201

Schwartz, M., Leikin, M. & Share, D. L. 2005. Bi-literate bilingualism versus mono-literate bilingualism: A longitudinal study of reading acquisition in Hebrew (L2) among Russian-speaking (L1) children. *Written Language and Literacy* 8: 179–207.

Silva-Corvalán, C. 1994. *Language Contact and Change. Spanish in Los Angeles*. Oxford: OUP.

Tsimpli, I. M. 2017. Multilingual education for multilingual speakers. *Open-access, policy paper in Languages, Society and Policy*. https://doi.org/10.17863/CAM.9803 <http://www.meits.org/policy-papers/paper/multilingual-education-for-multilingual-speakers>

Tsimpli, I. M. & Sorace, A. 2006. Differentiating Interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena. In *BUCLD 30: Proceedings of the 30th Annual Conference on Language Development*, Vol. 2, D. Bamman, T. Magnitskaia & C. Zaller (eds), 653–664. Somerville MA: Cascadilla.

Tsimpli, I. M., Sorace, A., Heycock, C. & Filiaci, F. 2004. First language attrition and syntactic subjects: a study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism* 8: 257–277. https://doi.org/10.1177/1367006904008003601

Unsworth, S. 2013. *UBILEC: Utrecht Bilingual Language Exposure Calculator*. Ms.

Vogindroukas, I., Protopapas, A. & Sideridis, G. 2009. *Expressive Vocabulary Assessment* (Δοκιμασία εκφραστικού λεξιλογίου). Chania: Glafki.

High sensitivity to conceptual cues in Turkish heritage speakers with dominant German L2

Comparing semantics–morphosyntax and pragmatics–morphosyntax interfaces

Elif Krause
University of Cologne

This study investigates semantics-morphosyntax and pragmatics-morphosyntax interfaces in separate experiments, yet examining the same structure, namely the optional verb number marking in Turkish. We tested a group of bilingual heritage speakers of Turkish, whose dominant language is German. Optional use of the overt verb number marking in Turkish interacts with semantic and pragmatic properties of the plural subject of the sentence. The interaction of optional verb number marking with these properties is tested separately in two different experiments, using the Magnitude Estimation technique. The results showed that the bilingual speakers treat both interface types differently from the monolingual speakers. More precisely, the bilingual speakers make finer distinctions regarding the semantic and pragmatic notions that were put into test. This sensitivity results in a semantically and pragmatically constrained pattern, which is in line with both language-specific descriptions and cross-linguistic tendencies. This outcome is taken to suggest that the nonconvergence in the bilingual data stems from a high sensitivity to the semantic and pragmatic properties that constrain the use of the morphosyntactic structure under investigation. The research results are further evaluated from a processing based linguistic framework, namely Modular Online Growth and Use of Language (MOGUL).

Keywords: heritage bilingualism, verb number marking, animacy, givenness, semantics-morphosyntax interface, pragmatics-morphosyntax interface, optionality, gradience

1.   Introduction

Earlier theoretical assumptions suggested a distinction between narrow syntax and syntax in the broad sense and assumed that each is operating on different grammatical values (e.g., Chomsky 1995; Jackendoff 2002). Data from agrammatic and child speech is in support of this distinction, suggesting that the latter demands higher processing resources (e.g., Platzack 1999, 2000; Avrutin 2004). Based on the theoretical assumptions on the internal organization of the modular linguistic system, earlier studies provided two types of accounts for non-target-like use of structures in child and agrammatic speech. The first one is a *language representation* account, which suggests that the representation of structures that require knowledge from more than one linguistic domain – such as syntax and semantics – is deficient in these speakers, i.e., normally developing children have difficulties in acquiring these structures, which emerge late in language development, and brain damage in agrammatic speakers distort exactly the same structures. The second one is a *language processing* account, and suggests that processing knowledge from more than one linguistic domain is more demanding and requires more processing resources. In this view, normally developing children and agrammatic speakers lack the capacity to perform these demanding processes since the child brain is not mature yet and the agrammatic speakers' brain is damaged.

Parallel to these observations, recent research in bilingualism has shown that structures relevant to the semantics-syntax interface and the pragmatics-syntax interface are more demanding to acquire, maintain and process when compared to narrow syntax. Following Platzack (1999), these studies predicted that the bilingual speakers have difficulties at the C-domain, at the pragmatic-syntax interface. Combining this idea with that of Döpke (1998), who showed that cross-linguistic influence (CLI) occurs on overlapping structures in the two languages of the bilinguals, these studies expected to see CLI on the partially overlapping structures at the interface between syntax and pragmatics. A series of studies reported non-native-like[1] comprehension and production in simultaneous and sequential bilingual acquisition (Hulk & Müller 2000; Müller & Hulk 2001; Paradis & Navarro 2003; Hacohen & Schaeffer 2007), L1 attriters (Gürel 2004; Tsimpli et al. 2004), near-native L2 speakers (Sorace 1999, 2000; Parodi & Tsimpli 2005) and adult early bilinguals (Polinsky 1997; Montrul 2005; Benmamoun et al. 2013). These studies addressed various phenomena such as CLI, language dominance and the quality and quantity of input. The findings have shown that interface structures are more problematic for bilingual speakers than purely syntactic structures. These studies

---

1.   Following Rothman & Treffers-Daller (2014) the term *(non-)monolingual-like* is used elsewhere throughout the paper.

adapted the earlier representation and processing accounts that have been originally put forward based on the outcomes of child and agrammatic speech, and evaluated these in the bilingual context. Thus, they attempted to explain the bilingual difficulties by referring to the deficient knowledge representation and/or lack of linguistic processing resources that were mentioned above.

Recent research further claimed the pragmatics-syntax interface to be more problematic than the semantics-syntax interface (Tsimpli & Sorace 2006; Sorace & Serratrice 2009). However, earlier studies focused mostly on the pragmatics-syntax interface, almost exclusively looking at the null-subject phenomenon of languages that pair with English. The few studies on the semantics-syntax interface, on the other hand, presented controversial evidence that shows problematic bilingual attainment at this interface as well (cf., Hopp 2004; Yuan 2010). Thus, there is need for more research on the semantics-syntax interface to shed more light on these mixed results. Furthermore, there is a crucial gap that calls for investigations on both interface domains, ideally comparing the two within the same experimental framework. Essentially, an examination of a hitherto unexplored structure and language pair is necessary in order to see whether previous accounts and claims on this linguistic interface are generalizable to comparable linguistic structures.

This paper takes optional verb number marking in Turkish as a test case to investigate the interface phenomenon in L1 speakers of Turkish whose dominant language is German.[2] This group characterizes as bilingual heritage speakers (HSs) as described in the literature (most recently, Kupisch & Rothman 2018). The use of the plural suffix on the verb is optional in sentences with a plural subject in Turkish. The overt occurrence of the suffix is constrained by various semantic-pragmatic properties of the plural subject in the sentence (e.g., Sezer 1978; Schroeder 1999). Optional verb number marking in Turkish enables establishing a direct comparison of the semantics-morphosyntax and the pragmatics-morphosyntax interface by keeping the structure constant. This structure is also comparable to those that have been investigated earlier (i.e. the null-subject phenomenon): It yields optionality by allowing one default zero option and one contextually distributed overt option where the use of overt number marking is constrained by the semantic and pragmatic properties of plural subjects. German obligatorily requires the overt marking of the verb number in sentences with plural subjects regardless of the semantic-pragmatic features of the plural subjects. Thus, the partial overlap at the linguistic interface, where one language provides one option (i.e., overt) and the other provides two (i.e., both overt and zero) presents an ideal test case to study the overlapping structures at the interface phenomenon. This study asks whether

---

2.   This paper provides a summary of the research presented in Bamyacı (2016) and further discusses its findings.

HSs perceive the well-formedness of sentences consisting of structures relevant to the semantics-morphosyntax and pragmatics-morphosyntax interfaces in a monolingual-like fashion and whether they have monolingual-like realization of the semantic (i.e., animacy) and the pragmatic (i.e., givenness) constraint on verb number marking in Turkish.

The research outcomes are further evaluated according to recent processing based approaches, in particular within Sharwood-Smith and Truscott's (2014) MOGUL framework. Based on previous theoretical assumptions on the modular mind (Fodor 1983) and linguistic modules (Chomsky 1995; Jackendoff 2002), MOGUL describes language acquisition, production and processing both in the monolingual and multilingual mind. By virtue of being a processing-based framework, MOGUL enables accounting for the patterns in the data from the bilingual speakers by referring to their language input and frequency of language use.

2. Study 1: Semantics-morphosyntax interface

2.1 Linguistic background: The interaction of animacy with number marking

The animacy level of entities interacts with number marking on noun phrases as well as on verbs across languages (Yamamoto 1999; Croft 1990; Haspelmath 2011). Although languages vary in the animacy categories they realize, animacy always splits the entities with regard to number marking in the same way: the higher the animacy level, the higher the likelihood for overt number marking (Comrie 1989: 188; Corbett 2000: 55). Animacy is shown to be a gradient linguistic notion and is open to intra-generational language change. Comrie (1989) has shown the instability of animacy as a constraint controlling the use of morphosyntactic structures across generations. Similarly, the optional verb number marking in Turkish may also give rise to variation across monolingual and bilingual speakers.

Although Turkish follows a regular paradigm whereby person and number features of the subjects are obligatorily expressed on the verb, overt verb number marking *-lAr* is optional in sentences with a plural subject in Turkish (Sezer 1978; Kornfilt 1997; Göksel & Kerslake 2005). Unlike sentences with inanimate subjects, which obligatorily take a verb with zero marking (1b), animate plural subjects are free to take a verb with or without overt number marking (1a, taken from Sezer 1978: 26):

(1) a.   Animate plural subject:
      Çilingir-ler   kapı-lar-I   aç-tı-Ø/-lar.
      locksmith.PL DOOR.PL-ACC open.PST-3SG/-3PL
      "Locksmiths opened the doors."

     b.   Inanimate plural subject:
         Anahtar-lar kapı-lar-ı   aç-tı-Ø/-*lar.
         key.PL   door.PL-ACC open.PST-3SG/*3PL
         "Keys opened the doors."

Bamyacı et al. (2014) provided the first set of psycholinguistic data regarding the interaction of animacy and optional verb number marking and established zero marking as the default in isolated sentences. In the acceptability judgment study they conducted, the zero marking received a significantly higher preference across all animacy categories, whereas the preference for the overt verb number marking showed gradient effects with an increased preference towards the higher ends of the animacy hierarchy (i.e., human > animal > inanimate). The first experiment in the current study replicates the experiment in Bamyacı et al. (2014) on a group of HSs and compares them to age and education matched monolingual speakers.

2.2   Experimental stimuli

This experiment solely focuses on the animacy factor on number marking. For that reason, the pragmatic context and specificity of subjects were kept constant across the experimental items. The experimental sentences consisted of a plural subject in the sentence initial position and a verb immediately following it.

    As shown below, the sentential subjects belong to four different animacy categories (32 sentences each), each further divided into two sub-categories on the basis of typological and language-specific observations (most recently see Haspelmath 2011). Although these sub-divisions yielded meaningful sub-levels for only the first three categories, the inanimate category was also divided into two further categories in order to maintain a balanced experimental design, see (2):

  (2)   Levels and sub-levels of the factor *animacy*:
      1.   Human:
          a.   Kinship (e.g., *anne* 'mother')
          b.   Profession (e.g., *öğretmen* 'teacher')
      2.   Animal:
          a.   High (e.g., *deve* 'camel')
          b.   Low (e.g., *örümcek* 'spider')
      3.   Quasi-animate:
          a.   Teleologically capable entities (e.g., *uçak* 'plane')
          b.   Entities with inherited animacy (e.g., *el* 'hand')
      4.   Inanimate:
          a.   Appliance (e.g., *sandalye* 'chair')
          b.   Clothes (e.g., *gömlek* 'shirt')

Each subcategory consisted of 16 sentences. Each sentence was used with an appropriate verb with and without overt number marking (*Kuş-lar öt-üyor-(lar)* [bird. PL sing.PROG-(PL)], "The birds are singing"), yielding 256 experimental items in total. The verbs were chosen among intransitive verbs in order to avoid any kind of influence an object might cause, such as bias towards individuated readings and length effects. Furthermore, in order to eliminate any bias towards an animate reading, two types of intransitive verbs, agent-subject verbs (*çalış-* 'to work') and theme-subject verbs (*üşü-* 'to be cold'), were used with an equal distribution across the experimental sentences.

64 sentences were added as fillers in order to broaden the range of the participants' judgment rates, e.g. to see the rating for an ungrammatical sentence. For this purpose, sentences with bare singular nouns belonging to human, animal or inanimate categories, and singular nouns that co-occurred with a quantifier were used. The animacy of the subjects in the filler sentences were also equally distributed across human, animal and inanimate categories. Each verb was used with zero or overt number marking. In both cases the use of overt verb number marking leads to ungrammaticality in Turkish, see (3) and (4). Zero (grammatical) and overt (ungrammatical) marking were distributed equally across the fillers.

(3) *Bilgisayar bozul-uyor-Ø/*-lar*
    computer break.down.PROG-3SG/*-PL
    "The computer is breaking down".

(4) *İki bilgisayar bozul-uyor-Ø/*-lar*
    two computer break.down.PROG-3SG/*-PL
    "The two computers are breaking down".

Altogether, the item list consisted of 320 sentences. The items were split into two lists, each having 192 items (128 experimental items plus 64 filler items; the filler items were kept constant across the lists). This way, each participant saw a given experimental item either with a verb marked or not marked for number. The stimuli were randomized for each participant.

2.3   Method

Considering the gradient nature of the animacy and givenness factors, which are separately investigated in this study, binary acceptability judgments would be inefficient in capturing various acceptability levels. Magnitude Estimation technique (ME), however, enables the speakers to freely report their preferences regarding linguistic stimuli. In an ME experiment, participants are asked to assign an arbitrary number greater than zero to a modulus stimulus. Participants can assign

any number to a stimulus, the only condition being that the number reflects the perceived ratio of acceptability of the current stimulus and the modulus stimulus. For instance, if the current stimulus is perceived as twice as good as the modulus stimulus, it should receive a number twice the value of the modulus stimulus. In this way, speakers can make as many (or as few) distinctions as they want. A modulus item, which evoked medium acceptability[3] and did not prime the use of overt or zero verb number marking, was used (*Kahraman-lar cesur-lar* [hero.PL courageous. PL], "Heroes are courageous"). ME technique is also advantageous in providing interval data, which allows for a more appropriate interpretation of the variance of the means, as well as the application of parametric statistical tests (e.g., Sorace & Keller 2005; Bard et al. 1996; Bader & Häussler 2010).

2.3.1   *Methods of data analysis*

As common in ME experiments, the participants used any number greater than zero, including decimals. Therefore, the data was converted into a common scale using z-transformation (see Bard et al. 1996). Data were analyzed using R (R Core Team 2012), including the R packages *lme4* (Bates et al. 2012) and *languageR* (Baayen 2008). The z-scores based on the raw ratings were subjected to analysis via the Linear Mixed Effects models (*lme* models). The factor (verb) number was constant across all the analyses with two values, namely overt marking and zero marking. The additional factors, such as animacy in the first study and givenness in the second, were added as fixed effects with their corresponding values. Participants and items were included in the statistical design as random effects (see Baayen 2008). Where appropriate, pairwise comparisons were conducted using t-tests with the Bonferroni correction (Westfall et al. 2011: 29).

2.4   Experiment I

2.4.1   *Participants*

18 monolingual speakers (12 female), aged between 22 and 29 ($M = 25$) were tested.[4] All participants were native speakers of Turkish, and none of them had started learning a second language before the age of 12. They were born in different regions of Turkey, but were living in Ankara at the time of testing. They had at least a high school degree, corresponding to a minimum of 11 years of formal education. They

---

3.   The modulus sentence was evaluated by several native speakers of Turkish who did not take part in the study before the experiments were conducted.

4.   The participants were recruited and tested by the author in all the experiments presented in this paper.

did not have any previous or continuing neurological disorders and had normal hearing and eyesight.[5]

2.4.2   *Results*

Figure 1 shows the mean z-scores for sentences with plural subjects belonging to each animacy category, from the highest animacy level to the lowest (Human, Animal, Quasi-animate, Inanimate). Mean ratings for each category are plotted separately for sentences with a plural verb (or overt verb number marking; blue bars) and for sentences with a singular verb (or zero verb number marking; purple bars). Positive z-scores indicate that the corresponding sentence type received ratings higher than the average rating; z-scores below zero indicate ratings below the average.
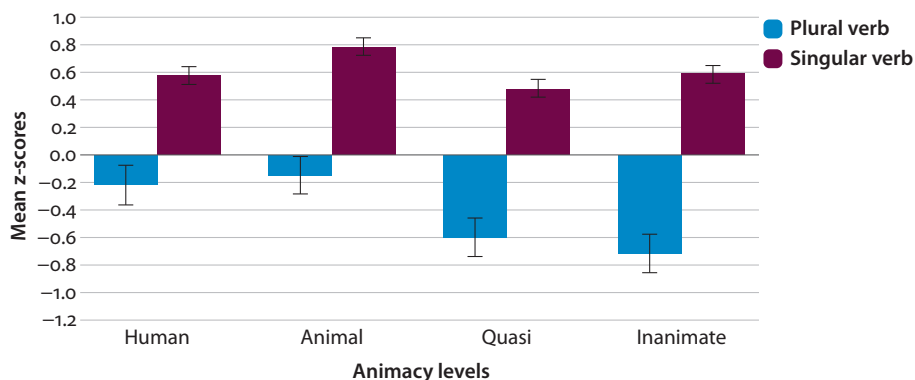
Figure 1. Interaction of verb number and animacy in monolinguals. Error bars represent +/− 2 standard errors (SE)[6]

This figure illustrates that the singular verbs are preferred significantly higher than the plural verbs across all animacy categories, justified by the pairwise comparisons using t-tests with Bonferroni correction (all *p*-values < .001).[7]

The preference for plural verbs decreases as the animacy level becomes lower, with a striking drop between the animate categories (Human and Animal) and

---

5.   The data from the monolingual speakers reported in this study are a sub-group of participants of an earlier study by Bamyacı et al. (2014). The speakers in this sub-group are age and education matched speakers to the bilingual speakers tested in experiment II presented in this paper.

6.   Error bars always represent +/− 2 standard errors (SE) across all the graphs throughout this paper.

7.   The pairwise t-test comparisons always included a Bonferroni correction procedure.

the inanimate categories (Quasi-Animate and Inanimate). There is a significant main effect of verb number in the *lme* analysis and an interaction effect for verb number and animacy when contrasting levels of Human and Animal, as well as the levels of Quasi-animate and Inanimate categories ($p < .05$ for both comparisons), see Table 1.

Table 1. *lme* results across main animacy levels for monolinguals

	Estimate	SE	t	p
(Intercept)	.08	.02	3.60	.000
VerbNr(Singular-Plural)	1.02	.03	32.56	.000
VerbNr*Animacy(Human-Animal)	.18	.08	2.06	.039
VerbNr*Animacy(Animal-Quasi-animate)	.11	.08	1.31	.191
VerbNr*Animacy(Quasi-animate-Inanimate)	.22	.08	2.52	.011

Overall, the monolingual speakers realize the following animacy categories:

Human > Animal | Quasi-Animate > Inanimate

A detailed analysis on sub-animacy levels show a constant preference for singular verbs confirmed by pairwise t-tests (all *p*-values < .001). However, there is no significant verb number – animacy interaction. In other words, the monolingual speakers did not differentiate any sub-animacy categories, see Table 2.

Table 2. *lme* results across main animacy levels in monolinguals

	Estimate	SE	t	p
(Intercept)	.14	.02	5.08	.000
VerbNr(Singular-Plural)	.94	.03	25.71	.000
VerbNr*Animacy(Kinship-Profession)	−.09	.12	−.74	.456
VerbNr*Animacy(High-Low animal)	−.03	.12	−.27	.784
VerbNr*Animacy(Teleological-Inherited)	.16	.12	1.27	.204

The analyses lead to three major results: First, speakers had a general preference for singular verbs, as in Bamyacı et al. (2014). Second, as expected, the speakers distinguished levels of animacy in a way that higher animacy increased the likelihood for overt marking. Unlike the default zero option, overt option is distributed across sentences depending on the animacy level of the subjects. Lastly, the monolinguals do not distinguish subanimacy levels.

2.5   Experiment II

2.5.1   *Participants*

HSs recruited for the two studies reported in this paper were born in Germany. Both of their parents were L1 Turkish speakers. They were exposed to only Turkish until about age 3. By this age, they attended German kindergartens and started being exposed to German. They were university students aged between 19 and 29 ($M = 23$) enrolled at various departments of a German university. None of them received formal education in Turkish. They did not have any previous or continuing neurological disorders and had normal hearing and eyesight.

25 (18 female) heritage speakers participated in the first experiment. The ages of the participants ranged between 19 and 29 ($M = 23$). The L1 Turkish is their weaker language and the majority language German is their dominant language. Their scores in the standardized language tests confirm this observation. The speakers had 75 % accuracy ($SD = 5.3$) in the Tömer Institute Turkish Language Proficiency Test,[8] and 95 % accuracy in the C1 level Language Proficiency Test of the Goethe Institute for German ($SD = 1.13$).[9]

2.5.2   *Results*

As Figure 2 shows, HSs portray a very similar pattern to that of monolingual speakers. The pairwise t-test comparisons across the animacy levels in both participant and item analysis confirmed the higher preference for singular verbs in the HS data ($p < .05$ for all categories).
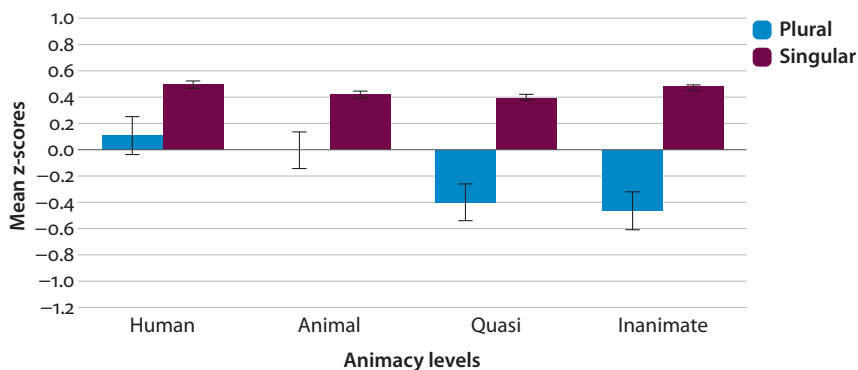
Figure 2. Interaction of verb number and animacy in HSs

---

8.   The Turkish proficiency test included questions on Turkish grammar as well as comprehension questions based on long paragraphs; the scores achieved in this test confirm that the heritage speakers have fair reading skills in Turkish.

9.   For a more detailed evaluation of these heritage speakers through data from questionnaires on language background and use as well as self-evaluations, see Bamyacı (2016).

Unlike in the monolingual data, however, there is a significant effect of verb number and a main effect of animacy, leading to a significant interaction between the two when Animal and Quasi-animate categories are compared (see Table 3). That is, the distinction between singular and plural verbs was sharper in the inanimate categories (Quasi-animate and Inanimate) compared to the animate categories (Human and Animal).

Table 3. *lme* results for main animacy levels in HSs

	Estimate	SE	t	p
(Intercept)	.12	.02	5.74	.000
VerbNr(Singular-Plural)	.65	.03	21.75	.000
VerbNr*Animacy(Human-Animal)	.04	.08	.55	.610
VerbNr*Animacy(Animal-Quasi-animate)	.37	.08	4.13	.000
VerbNr*Animacy(Quasi-animate-Inanimate)	.14	.08	1.66	.059

Unlike the ternary scale of the monolinguals, HSs made a two-way distinction: Human | Animal > Quasi-Animate | Inanimate.

HSs differed from the monolinguals in the detailed analysis of sub-animacy levels as well. Unlike the monolinguals, HSs did not carry over their tendency to have a higher preference for the singular verbs across all sub-animacy levels, see Figure 3. Pairwise t-tests show that the effect of verb number is not significant for the highest animacy categories (Kinship and Profession) in the participant analysis ($p < .05$). This result indicates that the realization of human entities is different from non-human entities, such as animals.
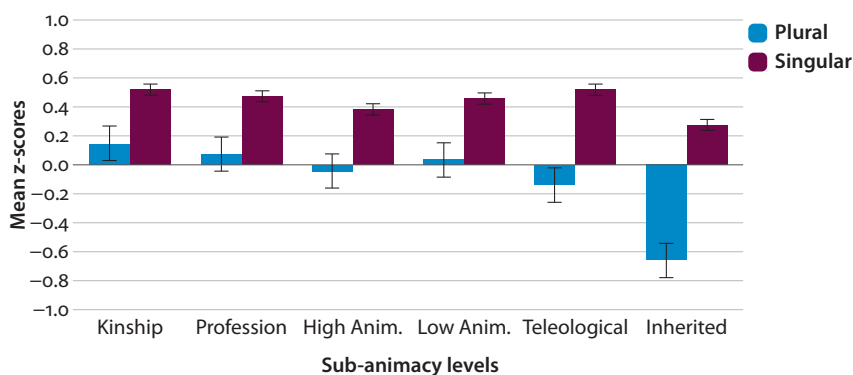
Figure 3. Interaction of verb number and sub-levels of animacy in HSs

The *lme* results revealed that unlike monolinguals, HSs were sensitive to the sub-levels of animacy. They had a stronger dispreference for the plural verbs when used with subjects referring to entities with inherited animacy, compared

to teleologically capable entities. This result carries the category of teleologically capable entities to a higher level than that of the category of entities with inherited animacy in the animacy model. This resulted in a significant interaction effect of verb number and animacy when the entities with teleological capacity were compared to the entities with inherited animacy, see Table 4.

Table 4. *lme* results for sub-animacy levels in HSs

	Estimate	SE	t	p
(Intercept)	.16	.02	6.95	.000
VerbNr(Singular-Plural)	.55	.03	15.68	.000
VerbNr*Animacy(Kinship-Profession)	−.03	.12	−.28	.777
VerbNr*Animacy(High-Low animal)	.01	.12	.08	.934
VerbNr*Animacy(Teleological-Inherited)	.27	.12	2.73	.023

2.6 Comparing the two groups

2.6.1 *Results*

Figure 4 shows a direct comparison of the ratings of the two groups, where the two groups seem to provide a very similar pattern. However, the distinction of the singular and plural verbs is sharper in the monolinguals when compared to HSs.
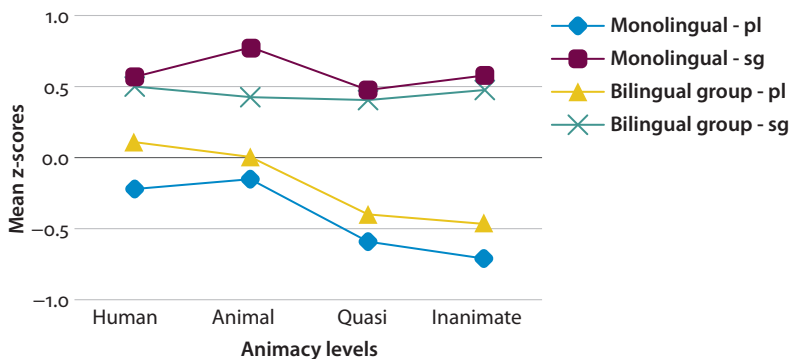
Figure 4. Mean z-scores of monolingual and HSs across main animacy categories

As explained above, monolinguals made a ternary distinction of the main animacy levels different from the two-level distinction of HSs, whereas the heightened sensitivity of HSs to the animacy levels became apparent in the detailed analysis of the sub-animacy levels. In contrast to the lack of distinctions of sub-animacy levels in the monolingual data, HSs differentiated sub-animacy levels in a semantically constrained fashion.

Table 5. *lme* results comparing the two group across main animacy levels

	Estimate	SE	t	p
(Intercept)	.106	.01	5.73	.000
Verb Number	.84	.02	37.63	.000
Group	−.04	.02	−2.05	.040
VerbNr*Group	.37	.04	8.58	.000
Group*VerbNr*Animacy(Human-Animal)	.10	.12	.86	.389
Group*VerbNr*Animacy(Animal-Quasi)	−.22	.12	−1.80	.071
Group*VerbNr*Animacy(Quasi-Inanimate)	.08	.12	.64	.519

These differences lead to a significant main effect of group and verb number and an interaction effect of the two ($p < .05$ for both, see Table 5). The group difference is particularly evident at the split between animates (Human – Animal) and inanimates (Quasi-animate – Inanimates).

2.7 Conclusions

Taken together, HSs behave non-monolingual-like at the semantics-morphosyntax interface. Given that German requires obligatory overt verb number marking, CLI would lead to overrating of the overt option in the HS data. But this was not the case. Thus, the HS pattern is free from CLI from the dominant language German. HSs rather portray a non-monolingual-like pattern, which is semantically constrained in line with the expectations of both grammar-specific descriptions of Turkish and of typological generalizations regarding the interaction of animacy and optional number marking across languages. Further implications will be discussed in the last section.

3. Study II: Pragmatics-morphosyntax interface

3.1 Linguistic background: The interaction of givenness with overt morphological marking

Various properties of the subject, such as referentiality, salience, agentivity, definiteness, distinctness or givenness in discourse have been suggested to interact with the verb number marking in Turkish (for a detailed overview see Bamyacı 2016). This study investigates the pragmatic properties that modulate the use of optional verb number marking. The animacy constraint is eliminated as this study looks solely on human plural subjects and investigates their occurrence in various pragmatic contexts.

Schroeder (1999) provides thus far the most elaborate evaluation on the interaction of verb number marking with discourse/pragmatic properties of human plural subjects. Schroeder (1999: 128–129) summarizes the factors that determine the appropriate use of singular or plural verbs under *distinctness motivation* and *pragmatic motivation*. The *distinctness motivation* requires the human referents of a plural subject to belong to a 'distinct' group. According to this view, besides being human and agentive, the shared knowledge between the speaker and the hearer triggers the use of overt verb number marking (Schroeder 1999: 128). For instance, the referents of plural subjects may belong to the same social group of the speaker as is the case in (5), where the speaker himself is a composer (Schroeder 1999: 126):

(5) *Besteci-ler   genellikle sözlü eser-ler   besteli-yor-lar.*
    composer.PL usually   vocal piece.PL compose.PROG-3PL
    "The composers usually compose vocal pieces."

It is important to point out that, besides conveying shared knowledge, being human and agentive, the referents of the plural subjects also act in the present or at a definite time period.

The *pragmatic motivation*, on the other hand, requires human plural subjects to be established as persistent topics. The human plural subjects that have been introduced earlier into the discourse, and are thus known to the hearer, gain a discourse topic status. An example to this is (6) where the plural subject *o kadar çok insan* 'that (or so) many people' is introduced into the discourse, and it is reintroduced with *insanlar* 'people' in the subject position. Schroeder suggests the second sentence (6b) to take a plural verb because of the earlier mention of its plural subject, see (6) taken from Schroeder (1999: 121):

(6) a. *Para-mız   var-ken   o   kadar çok   insan*
        Money.POSS.1PL exist.CONV that so   many person
        çevre-miz-de   fır dön-üyor-du-Ø   ki.
        Surrounding.POSS.1PL-LOC whirl around.PRES-PST.1SG PART
        "As long as we had money there were so many people whirling around us."
    b. *İnsan-lar sabah   kahvaltı-sı-nın   hangi saat-te*
        person.PL morning breakfast.POSS-GEN which time.LOC
        ol-du- ğu-nu   çok iyi   bili-yor-lar-dı.
        be.OP-POSS.3SG-ACC very good know.PROG-3PL-PST
        "People knew very well what time breakfast was at."

According to this motivation, the subjects of existential sentences do not take plural verbs, because they typically introduce new entities into the discourse. In a similar vein, interrogatives do not take a plural verb, because they are always treated as focus concepts.

Taken together, these two motivations direct us to understand that givenness is the factor that triggers the occurrence of overt verb number marking. That is, the information structural properties of the plural subjects, namely givenness and newness and the topic-focus structure, play a role on the use of zero and overt verb number marking.

In line with the discourse constraint described in this section, relative clauses (RC) typically involve a proposition, which conveys given information. By means of the information RC present, the entities modified by RC become identifiable (Givón 1990: 646). For instance, the subject in (7a) is suggested to be modified by a RC only if the speaker assumes that the hearer knows about the particular person being referred to by the subject; if not, the use of RC would be inappropriate and the sentence would look like (7b) (both examples taken from Haig 1997: 306):

(7) a. [The guy who phoned yesterday] called by this morning.
     b. A guy phoned yesterday and he called by this morning.

The characteristics of optional verb number marking in Turkish remind of the Switch Reference (SR) systems, where apart from their main functions, free or bound morphemes distinguish various phenomena, such as subject and non-subject (Hyman & Comrie 1981: x–xi), topic and non-topic (Foley & Van Valin 1984: 352; Chafe 1990: 60–61), actor and non-actor (Roberts 1997: 16), and agent and non-agent (Foley & Van Valin 1984: 119, 345; Roberts 1987: 292). Among many distinctions, the suggestion that "SR marks topicality" is the most widespread (Givón 1983; Dik 1997: 437; Ariel 1990: 138; Huang 2000: 287; Rising 1992: 4), and SR markers are suggested to keep track of the topic constituents. These cross-linguistic observations point to the commonalities between the phenomenon of optional verb number marking in Turkish and the SR systems.[10]

3.2   Experimental stimuli

The five experimental categories described below were defined in order to investigate the interplay between discourse-pragmatic factors and optional verb number marking.

Condition 1: Discourse topic
The first sentence introduced a human plural entity as a non-subject argument (8a) in order to avoid the expression of number on the verb so that neither overt nor zero marking is primed. The second sentence (8b) coherently expanded the context,

---

10. For a more detailed analysis see Bamyacı (2016).

without any reference to the human plural entities, which were mentioned earlier. The third sentence (8c), which was the critical one, reintroduced the human plural entities as the subjects of the sentence in sentence initial position.

(8) a. *Bu   haftasonu en   sev-diğ-im   yazar-lar-ın   imza*
         This weekend SUP love.REL-2SG author.PL-POSS.1SG autograph
         gün-ü   var.
         day.COMP exist.

   b. *Heyecan-la   bu-nu   bekli-yor-um.*
       Excitement-with this.ACC wait.PROG-1SG

   c. *Yazar-lar konuşma yap-acak-(lar).*
       Author.PL speech   do.FUT-**(3PL/-ø)**
       "My favorite authors have an autograph session this weekend. I am waiting for this in excitement. The authors will give a speech."

Condition 2: Modified topic
This category consisted of sentences with a human plural subject in sentence initial position modified by a relative clause, as in (9):

(9) *Zehir-len-en   hasta-lar hastane-de   yat-ıyor-(lar).*
     Poison.PASS-REL patient.PL hospital.LOC stay.PROG-**(3PL/-ø)**
     "Poisoned patients stay in the hospital."

Condition 3: Sentence topic
A bare plural subject referring to human entities was used sentence initially, as in (10).

(10) *Hasta-lar hastane-de   yat-ıyor-(lar).*
      patient.PL hospital.LOC stay.PROG-**(3PL/-ø)**
      "Patients stay in the hospital."

Condition 4: Focus
The first sentence presented a context without any human agents, as in (11a). In the second sentence (11b), which was the critical one, a constituent other than the subject of the sentence was topicalized and the human plural subject was presented in the preverbal position, which typically assigns focus status in Turkish (Erguvanlı 1984).

(11) a. *Duvar-ı boya-yacak kimse-yi   bul-ama-dı-k.*
          wall.ACC paint.FUT   nobody.ACC find.NEG-PST-1PL

    b. *Sonunda   duvar-ı   öğrenci-ler boya-dı-(lar).*
        In.the.end wall.ACC student.PL paint.PST-**(3PL/-ø)**
        "We couldn't find anybody to paint the wall. In the end the students painted the wall."

Condition 5: Focus interrogative

A plural question word[11] denoting a human plural entity is presented in an interrogative construction, as in (12). Regardless of their position in the sentence, the question words in interrogative constructions obligatorily have focus status in Turkish (Erguvanlı 1984). Thus, the occurrence of overt verb number marking is expected to be unlikely in such sentences.[12]

(12)   *Kim-ler kurabiye-yi ye-di-(ler)?*
       Who.PL cookie.ACC ate.PST-(3PL/-ø)?
       "Who ate the cookie?"

Overall, the plural human subjects presented in these experimental categories represent a hierarchy with regard to their identifiability. The identifiability depends on the criteria the subjects meet with regard to the factors that have been considered to motivate overt verb number marking (see Table 6 for a summary).

Table 6. Discourse properties of the human plural subjects used in each experimental condition

	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5
Discourse Topic	✓	✗	✗	✗	✗
Distinctness	✓	✓	✗	✗	✗
Sentence Topic	✓	✓	✓	✗	✗

Filler Type 1: Obligatory overt verb number marking

The first sentence consisted of a plural human subject (13a). The second sentence had a null subject that referred to the human referents of the overt subject in the first sentence (13b). Turkish requires obligatory verb number marking for sentences in which the subject is null. As the null subject refers to multiple human referents in this filler type, the overt verb number marking is obligatory (13b):

(13)   a.   *Filiz'in   kız-lar-ı   çok çalışkan.*
            Filiz'.POSS daughter.PL-ACC very hard.working
       b.   *Haftason-lar-ı   bile ders çalış-ıyor-lar/*Ø.*
            Weekend.PL-ACC even study work.PST-3PL/*ø
            "Filiz's daughters are very hard working. They study even on the weekends."

---

11. Question words can be pluralized in Turkish (Kornfilt 1997: 57).

12. Please note that an interaction between the ACC/ZERO marking on the direct object is not expected to result in a different judgment of the experimental items in this condition, because the use of overt verb number results in an ungrammatical construction regardless of the form in which the direct object occurs.

Filler Type 2: Obligatory zero verb number marking
The first sentence consisted of a singular human subject in the sentence initial position (14a). The second sentence had a null subject that referred to the human referent of the overt subject in the first sentence (14b). As the null subject referred to a singular human referent in this filler type, the zero verb number marking is obligatory (14b):

(14) a. *Filiz'in   kız-ı   çok çalışkan.*
       Filiz'.POSS daughter.ACC very hard.working
     b. *Haftason-lar-ı   bile ders çalış-ıyor-Ø/*lar.*
       Weekend.PL-ACC even study work.PROG-(Ø /*3PL)
       "Filiz's daughter is very hard working. She studies even on the weekends."

Overall, there were seven categories (five experimental and two filler conditions), each consisting of 14 items. Each sentence was used once with overt verb number marking and once without it. This doubled the number of items. A total of 196 items were first split into two lists. The two lists were created in different randomized orders and distributed equally among the participants.

3.3   Procedure

An experimental layout identical to the computer version of the first study was prepared using online forms, and the participants were asked to participate in the experiment by accessing these online forms at a quiet place. The participants were required to respond to the instructions and the practice session, in the same order as in the first study, without the option of skipping any of them.

On the main session, the modulus item consisting of two sentences, which evoked medium acceptability and did not prime the use of overt or zero verb number marking (*It's raining outside. Ali and his friends have their umbrellas with them*) was shown together with a fixed value of 40 on top of the page throughout the experiment. The participants could thus judge each critical item proportional to the modulus and assign values accordingly. Each critical item is presented once with and once without the overt verb number marking on the same page. Participants entered numerical values higher than 0 to each of the two sentences using the keyboard and pressed the enter button to see the next experimental item.

3.4   Experiment III

3.4.1   *Participants*
22 monolingual participants (15 female), aged between 20 and 29 ($M = 25.9$), who met the same criteria as in the first study, were tested.

3.4.2   *Results*
Figure 5 shows the mean z-scores for categories with plural subjects varying in their level of givenness, ordered according to the likelihood for the occurrence of overt verb number marking from the highest to the lowest.
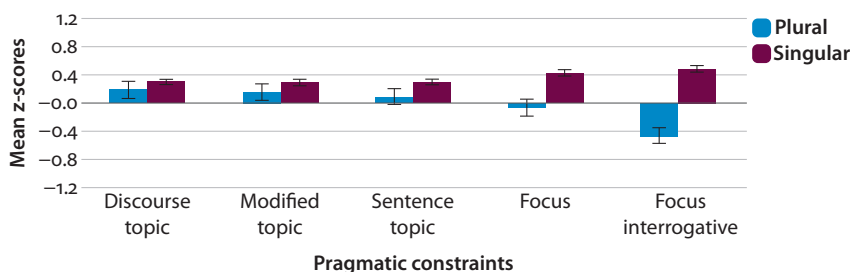
Figure 5. Interaction of verb number and givenness in monolinguals

The difference between the preference for plural and singular verbs was significant only for the Focus ($p < .05$) and the Interrogative conditions ($p < .001$) and not for the others, namely Discourse Topic, Modified Topic, and Sentence Topic (for all $p > .05$).

The results of the *lme* model showed a significant main effect of verb number as well as a significant interaction between verb number and constraint (i.e., givenness level) when the Sentence Topic (given) and Focus (new) conditions are compared.

Regardless of the degree of their givenness, all the given subjects received higher ratings both with singular and plural verbs. On the contrary, plural subjects with different levels of newness were distinguished from each other. The interaction of verb number and constraint is significant for the contrast between the Focus and Interrogative Focus, see Table 7. As expected, the distinction between singular and plural verbs is sharper in the Interrogative Focus condition than in the Focus condition.

Table 7. *lme* results for pragmatic constraints in monolinguals

	Estimate	SE	t	p
(Intercept)	.06	.01	4.12	.000
VerbNr(Singular-Plural)	.27	.03	8.14	.000
VerbNr*Constraints(DiscourseT.-ModifiedT.)	.01	.12	.10	.919
VerbNr*Constraints(ModifiedT.-SentenceT.)	−.08	.12	−.69	.487
VerbNr*Constraints(SentenceT.-Focus)	−.73	.12	−5.83	.000
VerbNr*Constraints(Focus-Interrogative F.)	−.44	.12	−3.48	.000

Overall, these results lead to the following scale, which indicates that, regardless of its strength, topicality (the given status) of human plural subjects increases the likelihood for the occurrence of overt marking, whereas focused plural subjects (with a new status) reduce it: Topic [Discourse Topic | Modified Topic | Sentence Topic] > Focus > Focus Interrogative.

3.5   Experiment IV

3.5.1   *Participants*
24 HSs (18 female), aged between 19 and 29 ($M = 22.3$), who met the criteria explained for the first study were tested.

3.5.2   *Results*
As Figure 6 illustrates HSs, like the monolinguals, showed a significantly higher preference for the Focus and Interrogative Focus conditions ($p < .001$), but not for the others ($p > .05$ for all)
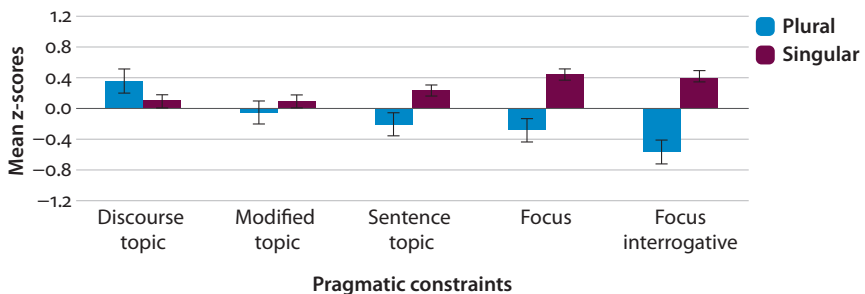
Figure 6. Interaction of verb number and pragmatic constraints in HSs

Interestingly, all the contrasts that compared the categories with each other in terms of to what extent the singular and plural verbs were distinguished from one another produced significant effects. In other words, HSs distinguished all the givenness and newness levels from each other with significance, see Table 8.

Table 8. *lme* results across pragmatics constraints in HSs

	Estimate	SE	t	p
(Intercept)	.02	.01	1.52	.127
VerbNr(Singular-Plural)	.33	.03	10.61	.000
VerbNr*Constraints(DiscourseT.-ModifiedT.)	.40	.11	3.39	.000
VerbNr*Constraints(ModifiedT.-SentenceT.)	−.28	.11	−2.42	.015
VerbNr*Constraints(SentenceT.-Focus)	−.53	.11	−4.51	.000
VerbNr*Constraints(Focus-InterrogativeF.)	−.25	.11	−2.12	.034

Overall, the analysis of the HS data leads to the scale below, whereby the preference for plural verbs decreases from left to right: Discourse Topic > Modified Topic > Sentence Topic > Focus > Interrogative.

3.6   Comparing the two groups

3.6.1   *Results*

The monolingual and HS data are presented in the line graphs below in Figure 7 and 8. Both graphs present similar tendencies towards a preference for singular and plural verbs in the conditions at the lower ends of the scale, namely Interrogative-Focus and Focus conditions. The two groups, however, differ in the givenness conditions. The monolingual data present very similar results for both singular and plural verbs, whereas the HS data show a graded increase for the overt number marking towards the categories in which the givenness level is higher.
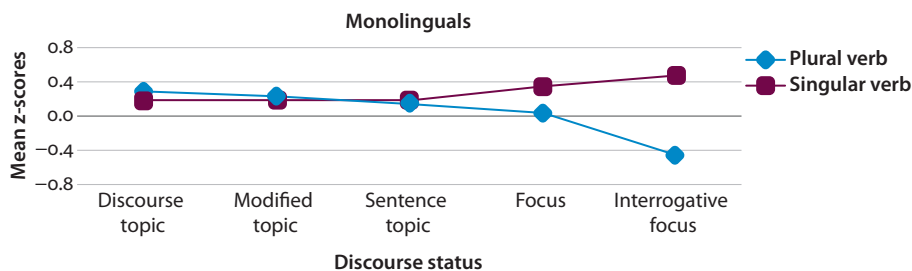
Figure 7. Preference for singular and plural verbs across pragmatic constraints in monolinguals
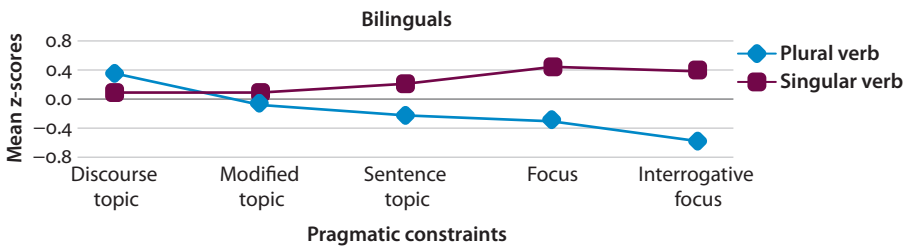
Figure 8. Preference for singular and plural verbs across pragmatic constraints in HSs

The *lme* model comparing the two groups shows a significant main group effect, see Table 9. The interaction of group and verb number is only significant when the Modified Topic and the Sentence Topic conditions are compared to each other.

Table 9. *lme* results comparing the two groups across pragmatics constraints

	Estimate	SE	t	P
Intercept	.04	.01	4.05	.000
Group	.04	.02	1.96	.049
Verb Nr	.30	.02	13.21	.000
Group*VerbNr(Singular-Plural)	−.06	.04	−1.31	.187
Group*VerbNr*Con.(DiscourseT.-ModifiedT.)	−.39	.17	−2.25	.024
Group*VerbNr*Con.(ModifiedT.-SentenceT.)	.19	.17	1.15	.249
Group*VerbNr*Con.(SentenceT.-Focus)	−.20	.17	−1.16	.242
Group*VerbNr*Con.(Focus-InterrogativeF.)	−.18	.17	−1.08	.276

3.6.2 *Conclusions*

Both monolingual and HSs differentiated the singular verbs from the plural verbs with significance only for the focused human plural subjects. This indicates that HSs realize the information structural status of subjects as a constraint on verb number marking in Turkish. The major difference between the two groups stems from the way they distinguish the levels of givenness. The monolingual speakers collapsed the three conditions in which the human plural subjects had different degrees of givenness into one level, whereas HSs showed sensitivity to the gradience in givenness and, unlike the monolinguals, made finer distinctions of the givenness levels. They reduced the strength of the distinction between the singular and plural verbs as the givenness level becomes weaker, following the expected trend: the higher the givenness, the higher the likelihood of the use of a plural verb. The results show that HSs behave non-monolingual-like at the pragmatics-morphosyntax interface. However, calling the HS pattern 'deficient', 'problematic', 'deviant', or using any label that implies inefficiency would be far-fetched. The givenness status of the

referents depends on various properties such as the identifiability of the referents of the arguments, the familiarity of the speaker and the hearer with the referents, the recoverability of their referents in a particular context. A variety of cognitively oriented proposals have attempted to explain the degrees of givenness or referentiality of arguments in a continuum, providing scales of predictability (Givón 1983: 55), givenness (Chafe 1994, 1976), familiarity (Prince 1981), or accessibility (Ariel 1990: 5). As the notion of givenness is gradient and not absolute, it is expected to find sensitivity to the levels of givenness. The HS pattern could rather be described as 'hyper-sensitive' or 'over-correct'. All in all, HSs behave non-monolingual-like at the pragmatics-morphosyntax interface, yet they portray a pragmatically constrained pattern.

4. Discussion

The two studies presented in this paper investigated the semantics-morphosyntax and pragmatics-morphosyntax interfaces in separate experiments, yet looking at the same structure, namely the optional verb number marking. The results of these experiments have shown that HSs behave non-monolingual-like at both interface types. More precisely, the nonconvergence in the HS data stems from a high sensitivity to the semantic and pragmatic properties that constrain the use of the morphosyntactic structure. This results in a semantically/pragmatically constrained pattern, which is in line with both language-specific descriptions and cross-linguistic tendencies regarding the structure under investigation. This section presents an evaluation of these outcomes in light of the previous literature.

Importantly, the differences between the monolinguals and HSs were on the more gradient areas of the interfaces, rather than less gradient areas. The use of overt verb number marking is not allowed with inanimate plural subjects in the first study, and it is not allowed with focused subjects in the second. These categories are arguably more categorical than the rest, because they require consistent rejection of the overt verb number marking. HSs successfully rejected the overt option in both studies at the more categorical experimental conditions, which do not yield optionality and therefore were less gradient. HSs, however, differed from the monolinguals on the gradient areas of the linguistic interfaces, which consist of categories requiring a gradient distribution of the two options comforting to the semantic-pragmatic constraints. In light of this data, it can be concluded that the phenomenon of linguistic interface is not problematic by itself for HSs. As the data suggests, HSs successfully handle the interface processing for more categorical areas within the interfaces; they coordinate the animacy information with the optional verb number marking like the monolingual speakers do for the inanimate category

in the first and the focus categories in the second study. When HSs are to distribute the overt option across categories that vary in their semantic/pragmatic properties, however, HSs provide finer distinctions thanks to their higher sensitivity to these properties. This sensitivity, on the other hand, systematically reflects the predictions of language-specific descriptions and cross-linguistic tendencies.

Similar to these outcomes, there is child speech data that reports non-adult-like yet semantically and pragmatically constrained patterns which signal sensitivity of the children to semantic-pragmatic properties. For instance, Allen (2000: 511) shows that the monolingual Inuktitut speaking children distribute the overt subject and object arguments across discourse contexts differently from the adults. A detailed data analysis shows that children use informative arguments significantly more often when compared to uninformative ones in line with the *principle of informativeness*, which suggests 'informativeness' to determine the use of overt arguments (proposed by Greenfield & Smith 1976). Allen takes this result to suggest an early (hyper)sensitivity to the pragmatic features of discourse (the dynamics of information flow in context) in overt or null argument selection in child speech (Allen 2000: 485). Similarly, Hoop and Krämer (2006) present results based on the interpretation of the referential status of subject and object arguments in children. Regardless of the constituent order or grammatical properties that may influence the referentiality of arguments, monolingual children speaking Dutch tend to perceive indefinite objects as non-referential, and indefinite subjects as referential (Hoop & Krämer 2006: 103). Interestingly, however, the non-target-like trend in argument interpretation of children accords with cross-linguistic observations (Hoop & Krämer 2006: 107). Cross-linguistically, the subjects tend to be referential, definite, topical, animate, and high prominent, whereas objects tend to be non-referential, indefinite, inanimate, and low prominent in the discourse (e.g., Aissen 2003; Comrie 1989; Lee 2003). In monolingual acquisition of Norwegian, children are observed to make fine distinctions in syntax and information structure from early on as well (see Westergaard 2009, 2011).

There is also data from various types of bilingual groups in which the bilingual speakers behave non-monolingual-like in processing of interface-syntax; yet their output is semantically-pragmatically constrained in accord with cross-linguistic tendencies. Serratrice et al. (2004) investigated the null-subject phenomenon and showed that an English-Italian bilingual child omitted the low informative arguments (e.g., first or second person, old, highly active, present in the physical context, not contrasted with another argument and unambiguous) significantly more often than the high informative ones (e.g., third person, new to the discourse, absent in the physical context, contrasted with another argument, or ambiguous) (Serratrice et al. 2004: 185). Similar to the monolingual children in Allen's (2000) study, the bilingual child was sensitive to pragmatic constraints in the use of structures related

to the syntax-pragmatics interface (Serratrice et al. 2004: 186). In another study Polinsky (2011) provide data on the use of relative clauses by HS of Russian living in the US. The non-monolingual-likeness found in the grammar of these speakers did not stem from L2 (English) word order knowledge (Polinsky 2011: 18). According to Polinsky "the gate between the two languages has been locked" (Polinsky 2011: 18). Consequently, this study suggested the absence of sustained input to be the cause of attrition (Polinsky 2011: 18), and further explained the trend in adult HS grammars in terms of generalizations based on typological observations. More precisely, the pattern in adult HS grammar, namely having access only to subject arguments for relativization, is in accord with suggestions of Keenan and Comrie (1977) regarding the universal constraint on relative clause formation: When a language limits its relativization to a subset of argument positions, it prefers relativizing the subject arguments. De Groot (2005), on the other hand, provides further support from a study that examined the differences between Hungarian spoken in Hungary (HH) and some Hungarian varieties spoken outside of Hungary (HO). He analyzed the non-monolingual-like patterns in HO grammars from a typological point of view and checked whether they are within the range of the generalizations and implicational hierarchies based on typological observations. Various structures relevant to different phenomena in HO were examined.[13] De Groot reported convergence to corresponding cross-linguistic tendencies in HO and concluded that the paths and processes of non-convergence in bilingual grammars can be captured in linguistic hierarchies based on typological observations and the implications of linguistic universals (de Groot 2005: 369). More recently, Kupisch (2014) provided evidence to the higher sensitivity of heritage speakers to semantic-pragmatic properties. She has shown that adult Italian heritage speakers with dominant German, opposite to the expectations, did not overuse the overlapping form in adjective placement in Italian, namely the postnominal position of adjectives. This outcome rejects a possible CLI from the dominant language. Kupisch takes this outcome to result from awareness about what is common and what is different in the two languages of the bilingual speakers, and she describes this pattern as *cross-linguistic overcorrection* (CLO) (Kupisch 2014: 231). She proposes that the bilinguals exaggerate the contrasts rather than the similarities between the two languages and explains the lack of CLI by proposing that the adult bilingual speakers are capable of easily demoting competing structures in the "other" language (Kupisch 2014: 232).

---

13. This included number marking on nouns and verbs, the use or absence of a copula in relational expressions, the use of overt pronouns, the loss of possessive markers on nouns, and the loss of object agreement, the use of analytic or disjunct expressions in the morphological domains of inflection, derivation, and compounding.

As the above examples clearly show, the two observations common to both studies presented in this paper, namely the lack of CLI and a semantically/pragmatically constrained pattern, find support from previous literature as well as current research into the grammar of monolingual children and various types of bilinguals. The discussion in the following section will evaluate the CLI-free pattern of HSs from a processing-based framework.

5. Evaluation of the outcomes from a processing-based linguistic framework and conclusions

Based on previous theoretical accounts on the internal organization of the linguistic system (e.g., Fodor 1983; Chomsky 1995; Jackendoff 2002), MOGUL provides a model that explains language processing, production, acquisition and maintenance in the monolingual and the multilingual mind. MOGUL differentiates a domain-specific core linguistic system, consisting of the two *domain-specific modules*: phonology and syntax (see Figure 9). These two systems are governed by UG principles, and the construction of grammatical components within these systems is often subsumed under the rubric of UG (Sharwood-Smith & Truscott 2014: 15, 22). As in the parallel architecture model (Jackendoff 2002), MOGUL considers semantics to be within the Conceptual Structure (CS). According to Jackendoff (2002), the conceptual structure contains more than just semantics, and this module is relevant to thought in general. Similarly, CS is outside of the core linguistic area in MOGUL, it is systematically linked to the situational context and the rest of cognition, and it is connected to the core linguistic system by means of an interface, which is itself *within* the core linguistic area.

When not linked to each other, these modules cannot accomplish a task. Each module has its own integrative processor which can only recognize and manipulate representations in its own particular code and is, at least in this sense, an *encapsulated module*. Consequently, MOGUL places the interface processors, shown as black arrows between the modules in Figure 9, that 'coindex' or 'register' these modules to each other. The interfaces that link and match the representations in adjacent stores are *bi-directional*. Beyond 'linking' the modules, interface modules cross-match the elements in adjacent modules (Sharwood-Smith & Truscott 2014: 20). Any stimulus turns into an input only after being processed in one of the modules (Sharwood-Smith & Truscott 2014: 36). The processing described here is crucially incremental. One module can start working before the other one concludes its own process.

During language processing, the processors pick the most appropriate items among many other available items that are competing with each other. MOGUL
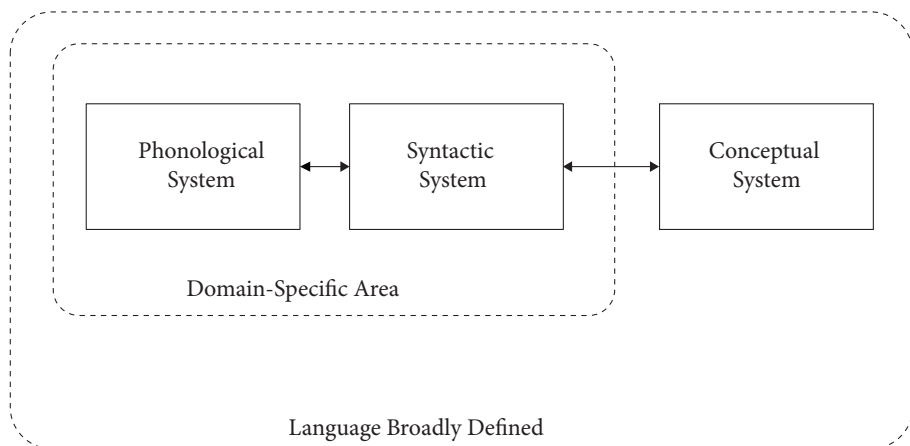
Figure 9. The language system in MOGUL
(adopted from Sharwood-Smith & Truscott 2014: 14)

bases the description of working memory on the activation levels of the items. Once an item (be it a lexical item or any other linguistic form) is used, its activation level increases and it becomes more readily available to participate in processing in the future (Sharwood-Smith & Truscott 2014: 69).

In MOGUL, language processing and language acquisition are closely related: The development of the language module simply results from the experience of language processing. MOGUL proposes the *Acquisition by Processing Theory* (APT), which characterizes language acquisition as "the lingering effects of processing". Here processing means the manipulation of items in the working memory, and acquisition means the transfer of these items into the long-term memory. First and second language acquisition processes operate according to these same principles, the only difference being that the latter occurs in the context of an existing language system, a natural consequence of which are additional complications in practice (Sharwood-Smith & Truscott 2014: 280).

In this framework, during processing of one of the languages in the bilingual mind, the related items in the other language will be active as well. This view suggests processing to be open to CLI, where the relevant candidates in both languages compete to be picked. Thus, high activation levels in L1 representations will inevitably exert an influence on the growth of a second language (Sharwood-Smith & Truscott 2014: 232). MOGUL resolves this problematic case by assuming the frequency of use to have a restricted power on changing the activation levels of items. Once a representation is created for an item, the activation level of the item rises sharply in the early stages, but, later, the curve levels off to the point at which additional exposure to input, no matter how extensive, has essentially no further

effect on the activation level of the item (Sharwood-Smith & Truscott 2014: 235). Thus, once the L2 representations are created, they are ready to participate in the competition during language processing without the need to reach a certain activation threshold.

The knowledge of the internal organization of the linguistic system and the way the modules function, enables us to explain the patterns in the data presented in this study. As both semantic and pragmatic knowledge is stored in the Conceptual Structure, and the interface module that links the CS to the Syntactic Structure is the same, it is expected to observe similar patterns in processing of both semantics-syntax and pragmatics-syntax interfaces. The lack of CLI finds an explanation as well. The linguistic structures do not need to reach a certain threshold to be able to participate in competition for being picked in language processing. Thus, the weak L1 structures have the chances to win the competition over the dominant L2 structures. Additionally, the HSs lifelong experience in controlling the two languages may enable them to separate the two languages from each other. MOGUL can also account for the gradient-nongradient distinction observed in the data. The categorical conditions, such as the inanimate and focused subjects, consistently demand zero verb number marking. Thanks to the robustness in these categorical conditions, the persistent processing experience of the bilingual speakers makes it possible for the interface module to create direct routes and to constantly reject the overt marking. Thus, the interface module in the bilingual mind is capable of coordinating semantic and morphosyntactic knowledge with efficiency. When, however, the input provides gradient data, where the overt option is allocated across certain semantic and pragmatic notions, because of the lack of systematic exposure to the language, the CS-SS interface module relies on its own mechanisms and provides patterns in accord with cross-linguistic tendencies, because this interface module is placed within the domain-specific core language area and functions in accord with UG. Therefore, as the CLI effects from the dominant language German are already excluded, the overt and the zero verb number marking in L1 keep competing in the bilingual mind and the processing results in a fine-grained semantically and pragmatically constrained pattern. Monolingual speakers, on the other hand, skip this competition thanks to their frequent language processing experience through interaction in the monolingual community, and as a result provide a parsimonious pattern, where their outcome seems to be less sensitive to the semantic and pragmatic constraints. In other words, the intensive exposure to data provides the monolingual speakers with the chances to create automatic and more economical routes in processing and production of gradient structures. In light of these data and the theoretical background, we can claim the optional structures to be steadily in competition with each other in the mind of HSs, which is a costly process. However, its outcomes are semantically and pragmatically constrained thanks to

the efficient functioning of the interface module that links the Conceptual Structure to the Syntactic Structure. More data from comparable structures, especially from online experiments that tap into the implicit knowledge of HSs, would be in order.

References

Aissen, J. 2003. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21: 435–483. https://doi.org/10.1023/A:1024109008573

Allen, S. 2000. A discourse-pragmatic explanation for argument representation in child Inuktitut. *Linguistics* 38(3): 483–521. https://doi.org/10.1515/ling.38.3.483

Ariel, M. 1990. *Accessing Noun Phrase Antecedents*. London: Routledge.

Avrutin, S. 2004. Optionality in child and aphasic speech. *Lingue e Linguaggio* 1: 67–89.

Baayen, R. H. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511801686

Bader, M. & Häussler, J. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46: 273–330. https://doi.org/10.1017/S0022226709990260

Bamyacı, E. 2016. *Competing Structures in the Bilingual Mind: A Psycholinguistic Investigation of Optional Verb Number Agreement*. Cham: Springer. https://doi.org/10.1007/978-3-319-22991-1

Bamyacı, E., Häussler, J. & Kabak, B. 2014. The interaction of animacy and number agreement: An experimental investigation. *Lingua* 148: 254–277. https://doi.org/10.1016/j.lingua.2014.06.005

Bard, E. G., Robertson, D. & Sorace, A. 1996. Magnitude estimation of linguistic acceptability. *Language* 72: 32–68. https://doi.org/10.2307/416793

Bates, D., Maechler, M. & Bolker, B. 2012. *lme4 Linear Mixed-Effects Models Using S4 Classes (R Package Version 0.999999–0)*. <http://cran.r-project.org/web/packages/lme4/index.html>

Benmamoun, E., Montrul, S. & Polinsky, M. 2013. Heritage languages and their speakers: opportunities and challenges for linguistics. *Theoretical Linguistics* 39(3–4): 129–181.

Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In *Subject and Topic*, C. N. Li (ed), 27–55. New York NY: Academic Press.

Chafe, W. 1990. Uses of the defocusing pronominal prefixes in Caddo. *Anthropological Linguistics* 32: 57–68.

Chafe, W. 1994. *Discourse, Consciousness and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago IL: University of Chicago Press.

Chomsky, N. 1995. *The Minimalist Program*. Cambridge: The MIT Press.

Comrie, B. 1989. *Language Universals and Linguistic Typology*. Chicago IL: University of Chicago.

Corbett, G. G. 2000. *Number*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139164344

Croft, W. 1990. *Typology and Universals*. Cambridge: CUP.

de Groot, C. 2005. The grammars of Hungarian outside Hungary from a linguistic-typological perspective. In *Hungarian Language Contact Outside Hungary: Studies on Hungarian as a Minority Language* [IMPACT: Studies in Language and Society 20], A. Fenyvesi (ed.), 351–370. Amsterdam: John Benjamins. https://doi.org/10.1075/impact.20.18gro

Dik, S. C. 1997. *The Theory of Functional Grammar: Complex and Derived Constructions*. Berlin: De Gruyter. https://doi.org/10.1515/9783110218374

Döpke, S. 1998. Competing language structures: The acquisition of verb placement by bilingual German-English children. *Child Language* 25: 555–584.
https://doi.org/10.1017/S0305000998003584

Erguvanlı, E. E. 1984. *The Function of Word Order in Turkish Grammar*. Berkeley CA: University of California Press.

Fodor, J. A. 1983. *The Modularity of Mind*. Cambridge MA: The MIT Press.

Foley, W. A. & Van Valin, R. D. 1984. *Functional Syntax and Universal Grammar*. Cambridge: CUP.

Givón, T. 1983. Topic continuity in discourse: The functional domain of switch-reference. In *Switch-Reference and Universal Grammar* [Typological Studies in Language 2], J. Haiman & P. Munro (eds), 51–83. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.2.06giv

Givón, T. 1990. *Syntax: A Functional-Typological Introduction*, 2nd ed. Amsterdam: John Benjamins.

Göksel, A. & Kerslake, C. 2005. *Turkish: A Comprehensive Grammar*. London: Routledge. https://doi.org/10.4324/9780203340769

Greenfield, P. & Smith, J. H. 1976. *The Structure of Communication in Early Language Development*. New York NY: Academic Press.

Gürel, A. 2004. Selectivity in L2-induced L1 attrition: a psycholinguistic account. *Journal of Neurolinguistics* 17: 53–78. https://doi.org/10.1016/S0911-6044(03)00054-X

Hacohen, A. & Schaeffer, J. 2007. Subject realization in early Hebrew-English bilingual acquisition: The role of cross-linguistic influence. *Bilingualism: Language and Cognition* 10(3): 333–344. https://doi.org/10.1017/S1366728907003100

Haig, G. 1997. On some strategies for case recovery in Turkish relativization. In *The Mainz Meeting: Proceedings of the Seventh International Conference on Turkish Linguistics, August 3 – 6, 1994*, L. Johanson (ed.), 299–320. Wiesbaden: Harrassowitz.

Haspelmath, M. 2011. Occurrence of nominal plurality. In *The World Atlas of Language Structures*, M. Haspelmath (ed), 142–145. Oxford: OUP.

Hoop, H. & Krämer, I. 2006. Children's optimal interpretations of indefinite subjects and objects. *Language Acquisition* 13(2): 103–123. https://doi.org/10.1207/s15327817la1302_4

Hopp, H. 2004. Syntax and its interfaces in L2 grammars: situating L1 effects. In *Proceedings of GALA (Generative Approaches to Language Acquisition) 2003* [LOT Occasional Series 3], J. van Kampen & S. Baauw (eds), 211–222. Utrecht: Utrecht University.

Huang, Y. 2000. *Anaphora: A Cross-Linguistic Approach*. Oxford: OUP.

Hulk, A. & Müller, N. 2000. Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition* 3(3): 227–244. https://doi.org/10.1017/S1366728900000353

Hyman, L. & Comrie, B. 1981. Logophoric reference in Gokana. *Journal of African Languages and Linguistics* 3: 19–37. https://doi.org/10.1515/jall.1981.3.1.19

Jackendoff, R. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780198270126.001.0001

Keenan, E. L. & Comrie, B. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8: 63–99.

Kornfilt, J. 1997. *Turkish*. London: Routledge.

Kupisch, T. 2014. Adjective placement in simultaneous bilinguals (German-Italian) and the concept of cross-linguistic overcorrection. *Bilingualism: Language and Cognition* 17: 222–233. https://doi.org/10.1017/S1366728913000382

Kupisch, T. & Rothman, J. 2018. Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism* 22(5): 564–582.

Lee, M. 2003. Dissociations among functional categories in Korean agrammatism. *Brain and Language* 84: 170–188. https://doi.org/10.1016/S0093-934X(02)00515-1

Montrul, S. 2005. Second language acquisition and first language loss in adult early bilinguals: exploring some differences and similarities. *Second Language Research* 21(3): 199–249. https://doi.org/10.1191/0267658305sr247oa

Müller, N. & Hulk, A. 2001. Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition* 4(1): 1–21. https://doi.org/10.1017/S1366728901000116

Paradis, J. & Navarro, S. 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): 1–23. https://doi.org/10.1017/S0305000903005609

Parodi, T. & Tsimpli, I. 2005. 'Real' and apparent optionality in second language grammars: Finiteness and pronouns in null operator structures. *Second Language Research* 21(3): 250–285. https://doi.org/10.1191/0267658305sr248oa

Platzack, C. 1999. The vulnerable C-domain. *Brain and Language* 77: 364–377. https://doi.org/10.1006/brln.2000.2408

Platzack, C. 2000. Multiple interfaces. In *Cognitive Interfaces: Constraints on Linking Cognitive Information*, E. van der Zee & U. Nikanne (eds), 21–53. Oxford: OUP.

Polinsky, M. 1997. Cross-linguistic parallels in language loss. *Southwest Journal of Linguistics* 14(1–2): 87–123.

Polinsky, M. 2011. Reanalysis in adult heritage language: A case for attrition. *Studies in Second Language Acquisition* 33: 1–24. https://doi.org/10.1017/S027226311000077X

Prince, E. F. 1981. Towards a taxonomy of given-new information. In *Radical Pragmatics*, P. Cole (ed.), 223–233. New York NY: Academic Press.

R Core Team. 2012. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rising, D. P. 1992. *Switch Reference in Koasati Discourse*. Dallas TX: Summer Institute of Linguistics.

Roberts, J. R. 1987. *Amele*. London: Croom Helm.

Roberts, J. R. 1997. Switch-reference in Papua New Guinea: A preliminary survey. In *Papers in Papuan Linguistics 3*, A. Pawley (ed.), 101–241. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Rothman, J. & Treffers-Daller, J. 2014. A prolegomenon to the construct of the native speaker: heritage speaker bilinguals are natives too! *Applied Linguistics* 35(1): 93–98. https://doi.org/10.1093/applin/amt049

Schroeder, C. 1999. *The Turkish Nominal Phrase in Spoken Discourse*. Wiesbaden: Harrassowitz.

Serratrice, L., Sorace, A. & Paoli, S. 2004. Subjects and objects in English-Italian bilingual and monolingual acquisition. *Bilingualism: Language and Cognition* 7(3): 183–205. https://doi.org/10.1017/S1366728904001610

Sezer, E. 1978. Eylemlerin çoğul öznelere uyumu. *Genel Dilbilim Dergisi* 1: 25–32.

Sharwood-Smith, M. & Truscott, J. 2014. *The Multilingual Mind: A Modular Processing Perspective*. Cambridge: CUP.

Sorace, A. 1999. Initial states, end states and residual optionality in L2 acquisition. In *Proceedings of the 23rd Annual Boston University Conference on Language Development*, A. Greenhill, H. Littlefield & C. Tano (eds), 666–674. Sommerville MA: Cascadilla Press.

Sorace, A. 2000. Syntactic optionality in non-native grammars. *Second Language Research* 16(2): 93–102. https://doi.org/10.1191/026765800670666032

Sorace, A. & Keller, F. 2005. Gradience in linguistic data. *Lingua* 115(11): 1497–1524. https://doi.org/10.1016/j.lingua.2004.07.002

Sorace, A. & Serratrice, L. 2009. Internal and external interfaces in bilingual language development: beyond structural overlap. *International Journal of Bilingualism* 13: 195–210. https://doi.org/10.1177/1367006909339810

Tsimpli, I. & Sorace, A. 2006. Differentiating interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena. In *Proceedings of the 30th Annual Boston University Conference on Language Development*, Vol. 2, D. Bamman, T. Magnitskaia & C. Zaller (eds), 653–664. Somerville MA: Cascadilla Press.

Tsimpli, I., Sorace, A., Heycock, C. & Filiaci, F. 2004. First language attrition and syntactic subjects: a study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism* 8(3): 257–277. https://doi.org/10.1177/13670069040080030601

Westergaard, M. 2009. *The Acquisition of Word Order: Micro-cues, Information Structure and Economy* [Linguistik Aktuell/Linguistics Today 145]. Amsterdam: John Benjamins. https://doi.org/10.1075/la.145

Westergaard, M. 2011. Subject positions and information structure: The effect of frequency on acquisition and change. *Studia Linguistica* 3: 299–332. https://doi.org/10.1111/j.1467-9582.2011.01183.x

Westfall, P., Tobias, R. & Wolfinger, R. 2011. *Multiple Comparisons and Multiple Tests Using SAS*, 2nd edn. Cary NC: SAS Publishing.

Yamamoto, M. 1999. *Animacy and Reference: A Cognitive Approach to Corpus Linguistics* [Studies in Language Companion Series 46]. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.46

Yuan, B. 2010. Domain-wide or variable-dependent vulnerability of the semantics-syntax interface in L2 acquisition? Evidence from wh-words used as existential polarity words in L2 Chinese grammars. *Second Language Research* 26(2): 219–260. https://doi.org/10.1177/0267658309349421

The Frequency Code and gendered attrition and acquisition in the German-English heritage language community in Vancouver, Canada

Esther de Leeuw

Queen Mary University of London

The paper investigates pitch level and span in a group of German L1-English L2 late bilinguals in comparison to two monolingual control groups. The late bilinguals had moved to Vancouver, Canada in adulthood, and had been living in Vancouver for an average of 40 years. The results indicate that the bilingual males increased their pitch in both English and German, and widened their pitch span, therefore indexing non-aggressive, friendly behaviour, but deviating from both monolingual pitch norms. Thus, the results offer evidence that pitch changes are at least in part dependent on the social and political environment in which they are embedded, as a low pitch level is associated with dominance and aggression which would boost the negative image of the Vancouver German community due to their ethnic origin after WWII.

Keywords: German, L1 attrition, prosody, pitch, Frequency Code, social significance

1. Introduction: First language attrition as it relates to heritage language research

Research into first language (L1) attrition usually examines bilinguals in a second language (L2) environment, in order to determine whether the native language has changed (i.e. undergone attrition) through acquisition of the L2. Comparing the two fields, in research investigating L1 attrition (henceforth attrition) within the domain of phonetics, subjects are most often *late* sequential bilinguals, who are investigated as adults (see, e.g., Cherciov 2013; de Leeuw 2019; Opitz 2013; Schmid 2011; Schmid & Köpke 2007 for a review on some attritional research). In contrast, in heritage language research, subjects are usually *early* sequential bilinguals, who are investigated either as children or as adults (Kupisch & Rothman 2018; Montrul 2008; Polinsky 2008; Rothman 2007; Rothman & Treffers-Daller 2014). However,

the suggestion put forth here is that individuals who have undergone attrition of their native language are comparable to heritage language speakers, and that therefore the two fields of research in linguistics are inextricably linked.

For example, Polinsky (2008: 40) has defined a heritage speaker of language x as "an individual who grew up speaking (or only hearing) [that language] as [his or her] first language but for whom [language x] was then replaced by another language as dominant and primary". According to this definition, the subjects of the current investigation, who grew up speaking German, which was largely replaced by English when they moved to Vancouver, Canada, are comparable to heritage language speakers, as they have potentially undergone attrition and experienced a shift in the dominant language. Both Montrul (2008) and Polinsky (2008) emphasise that heritage speakers are not a homogeneous group, but rather form a spectrum between those who may only have overheard language x through e.g. caregivers or before adoption (Au, Knightly, Jun & Oh 2002; Bowers, Mattys & Gage 2009; Pallier et al. 2003) to highly proficient speakers who may only lack certain registers in their language. Similarly, Kagan and Dillon (2001) refer to Russian Americans who moved to the United States from Russia in middle school as heritage speakers of Russian. As such, it is possible to view the late sequential bilinguals, investigated within the realms of attrition here, as native speakers of German who are similar to heritage language speakers (see also e.g. Rothman & Treffers-Daller 2014). In the current study, they form a crucial component of the German heritage language community in Vancouver, Canada, where they have in most cases lived for the majority of their lives and integrated to greater or lesser extents into this community.

The results from this study into acquisition and attrition of pitch therefore directly feed into our knowledge of the German heritage language community in Canada, which is a large immigrant population, having kept a low profile largely for historical reasons. In the 2006 Canadian Census (which was the closest Census before the time of the data collection), German was the third most common non-official mother tongue in Canada with 466,650 Canadians who described German to be their mother tongue. According to the Canadian Census, of those Canadians with German as mother tongue, approximately 179,000 immigrated to Canada before 1981 whilst even more Canadians who listed German as a mother tongue were born in Canada: 221,000. Between 1981 and 2001, 46,000 persons with German as a mother tongue arrived in Canada and 15,000 arrived in the five years preceding the 2006 Census. This data from the Canadian Census suggest that new Germans continue to arrive in Canada regularly, and that they are passing German on to their children, who often go on to describe German as their own mother tongue.[1]

---

1.   These data were accessed from StatsCan on 15 January 2017 <http://www12.statcan.gc.ca/census-recensement/2006/as-sa/97-555/figures/c1-eng.cfm>

More specifically, in Vancouver, British Columbia (BC), where the data for this project were elicited, there are also concentrated German "pockets". For example, in 2011, the local city newspaper, The Vancouver Sun, mapped ethnicities in Vancouver through an online tool and reported that people of German background made up the largest white ethnic group in Vancouver after those with a connection to the United Kingdom (Todd 2011). The newspaper reported that within the Vancouver area (i.e. Metro Vancouver), people with German background were highly concentrated around the community of Aldergrove, where they made up 22 per cent of the population, and that there was a similar pocket of German Canadians in the Port Kells region of northern Surrey. However, in the same online report (Todd 2011: § 44), it was documented that most people with German origins in Metro Vancouver tend "to keep a low profile about their ethnic origin", including "some of the more than 400,000 who moved to Canada after WWII, when some people of German background were still associated with the 'enemy'".[2]

Along these lines, it has been reported that German nationals who immigrated to Canada were perceived to have an "immediate connection with the Hitler dictatorship and likely felt some guilt or shame" and were therefore "generally less inclined to lead [German] organizations [in BC], if they joined them at all". For example, Leo K. documented being called a Nazi after moving to Vancouver:

> It did two things for me, really. It made me want to learn how to speak English very quickly to avoid being identified as German and it gave me the ability to run very fast. I was beaten up because I was a Kraut.   (Lieb 2008: 310)

Indeed, some Germans encountered physical and mental abuse for being German, being told at a job interview, for example, that they simply did not hire Germans at managerial level (Lieb 2008: 310). As a result of the Nazi-German history, Germans in BC, and most likely throughout Canada, have tended to play down attributes associated with "Germanness", e.g. "I was always afraid of people's reactions when they found out, though my family never had anything to do with the Nazis" (Lieb 2008: 307). At the same time, however, numerous quiet initiatives have encouraged the German language and culture to be maintained throughout Canada, and Germans have often continued to pass on the German heritage to their children (Lieb 2008). Such maintenance initiatives nevertheless elucidate how over 220,000 Canadians list German as their mother tongue, although born in Canada, and indicate that the German heritage language community is discrete, yet enduring.

---

2.   This newspaper article was accessed from The Vancouver Sun on 15 January 2017 <http://vancouversun.com/news/staff-blogs/ethnic-mapping-5-find-metros-dutch-blacks-germans-and-iranians>

Therefore, of the 26,935 German mother tongue speakers in Vancouver, as documented by the 2011 Canadian Census,[3] and the 84,605 German home language speakers in BC, as documented by the 2001 Canadian Census,[4] it is possible to infer that although this heritage language community might keep a lower profile in comparison to other large minority groups in Canada, many are still passing their native language on to their own children through encouraging their children to speak German at home and to attend German language Saturday schools (Lieb 2008).

Accordingly, if the late sequential German-English bilinguals of this study evidence attrition, it is possible that this will influence the transmission of German to subsequent generations in the German language heritage community in Vancouver. Specifically, if the late German-English bilinguals in the current investigation evidence attrition of pitch, there is potential to infer that their "attrited" German prosody will feed into the overall linguistic profile of the German community in Vancouver, Canada (see also Dominguez & Hicks 2016).

By looking at the "source", i.e. the adult generation of immigrants, as is done here, it is possible to investigate the question of what is present to be acquired in later generations. If some of "the missing properties [are not] available to the heritage language learner to acquire in the first place" (Rothman 2007: 385) as a result of attrition in the adult generation, heritage language learners would never be able to acquire that which was not available to them. Therefore, this research feeds more broadly into heritage language research by examining attrition and acquisition of pitch in German-English late bilinguals in the adult generation of the German-English bilingual community in Vancouver, Canada. How do German native speakers in Vancouver, who immigrated as adults to Canada, go about realizing pitch in their new language of English, and how does this affect their native language, if at all?

The German-English bilingual females and males have contrasting acquisitional tasks, assuming the target is the monolingual English norm, as will be explained in more detail below. Do the German-English bilinguals acquire the gender specific pitch norms in their new language, and do these new gender specific pitch norms affect their native language German? These questions will be explored in the present chapter, with findings related to the overall desire on the part of the

---

3. These data were accessed from StatsCan on 30 November 2017 <http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo12g-eng.htm>

4. These data were accessed from StatsCan on 30 November 2017 <http://www12.statcan.ca/english/census01/products/standard/themes/Rpeng.cfm?TABID=2&LANG=E&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=0&GK=0&GRP=1&PID=55542&PRID=0&PTYPE=55430,53293,55440,55496,71090&S=0&SHOWALL=0&SUB=0&Temporal=2001&THEME=41&VID=0&VNAMEE=&VNAMEF=>

German-Canadians to keep a relatively low profile, and disassociate from any potential connection with the Hitler dictatorship (Lieb 2008), and therefore, with being the "enemy" (Todd 2011: § 44). Accordingly, the Frequency Code (Gussenhoven 2004; Ohala 1983, 1984) which, very generally, proposes that lower frequencies are associated with dominance and aggression, is particularly relevant to this investigation, as it is these characteristics which in part characterise the Hitler dictatorship, and hence the "enemy".

2.   Purpose of this study

The primary objective of this study was to conduct a pitch analysis of attrition in female and male German-English late sequential bilinguals in Vancouver, Canada, and embed these findings into our knowledge of the German heritage language community in BC, Canada. This undertaking is interesting for two reasons.

Firstly, the potential attrition is hypothesized to occur differently in female and male late bilinguals in Canada due to the different language norms for female and male pitch in German and English. It is therefore also important to emphasize the social dimension underpinning this research, i.e. that languages are transmitted not entirely as homogeneous systems. Instead, the tasks of the German-English late bilingual females and males are different, dependent upon their gender.[5]

With regard to monolinguals, sociolinguistic research has shown that speakers use the variation inherent in all languages to different extents and for a range of different purposes (Cheshire, Kerswill, Fox & Torgersen 2011; Eckert & Rickford 2001; Labov 1963; Tagliamonte & D'Arcy 2009; Trudgill 1986). Specifically, with regard to pitch in monolinguals, a discourse analysis of 'Heath', an openly gay American English native speaker male, showed that in conversation with his friends, he employed higher pitch frequencies for longer durations, potentially to construct a 'diva' persona (Podesva 2007). Moreover, it has been found that in the speech of a group of Israeli men, who were all involved in gay activism, mean pitch levels were significantly higher and pitch span was significantly wider for gay topics than for non-gay topics (Levon 2009). It may be that the significant differences between gay and non-gay topics are "evidence of the men's desire to initiatively portray distinct 'gay' and 'non-gay' selves" (Levon 2009: 52). Such findings suggest that pitch may vary in males, potentially dependent on their sexual orientation, due to both the conversational setting, as well as the topic at hand. Moreover, as in the latter study

---

5.   Here, I use the term 'gender' to incorporate social and cultural differences, i.e. assuming similar anatomy of German and English males, and similar anatomy of German and English females.

the subjects were political activists, it may be that political affiliations, or disaffiliations, prompt changes in pitch patterns, which is of relevance to the study at hand. Here, it is proposed that the German-English bilinguals in Canada may want to dissociate from any "connection with the Hitler dictatorship", as is often reported by members of the German heritage community in Canada (Lieb 2008: 310).

That said, we know very little about how bilinguals, who use more than one language on a regular basis, acquire the varieties of their new language, and potentially attrite dependent on these varieties and their social meaning. Therefore, for methodological reasons, this study is of interest to research in bilingual communities more generally, as it takes a step forward to encourage investigations which will examine how within language variation is navigated and produced by bilingual communities.

An additional objective of this research was to relate the findings on attrition to research on heritage language acquisition, largely because the bilinguals in the present study are indeed considered to be heritage language speakers, embedded within their own unique community in Vancouver, Canada. Moreover, it is possible that their production of pitch might potentially influence what is able to be acquired by subsequent generations in the German-English bilingual community, and how this community as a whole develops over time, as time passes since WWII.

3.   Defining pitch

Pitch is a perceptual property, whereas fundamental frequency (F0) is generally investigated as its acoustic correlate (Hewlett & Beck 2006). At the articulatory level, F0 represents the number of vocal fold vibrations per second (Hewlett & Beck 2006). Auditory filters in the ear make frequency resolution become poorer as frequency increases. Despite this difference between pitch and F0, the two terms are often used synonymously and unless otherwise specified this will be the case here.

Pitch range, as a percept, has been further characterized using the terminology of pitch level and pitch span (Ladd 2008). Pitch level is defined as the overall pitch height whereas pitch span reflects how much pitch varies within a given speech sample (Mennen 2007; Mennen, Schaeffler & Docherty 2012). The mean of F0 has been used as the acoustic correlate of pitch level, i.e. expressed in Hertz (Hz), whilst pitch span is generally expressed in semitones (ST) (Mennen, Schaeffler & Docherty 2012; Patterson 2000), therefore normalizing pitch span.

Speaker specific anatomical differences impact an individual's pitch level, e.g. males generally have thicker and longer vocal folds than females, which result in a lower F0, i.e. lower pitch level (Hayward 2000; Laver 1980; Neppert 1999). Moreover, pitch is affected by the processes of aging (Linville 1996; Nishio & Niimi

2008). Male speaking F0 "lowers from young adulthood into middle age and then rises again into old age" (Linville 1996: 191), with male F0 rising from approximately 110 Hz at age 30 to approximately 130 Hz at age 70. In females, on the other hand, speaking F0 "appears to remain fairly constant until menopause when a drop in fundamental frequency (F0) occurs" (Linville 1996: 191). However, as will be discussed in the participant section, age was controlled in this study, and, therefore, any potential aging effects would have occurred similarly across all groups. It is most important to emphasize here that although pitch varies according to anatomical differences, speakers do habitually adopt a pitch range (Laver 1980), which may result in changes in the bilinguals of the current study, as was the case in the monolingual males in Podesva (2007) and Levon (2009), as discussed in more detail below.

4.   The Frequency Code

According to Ohala (1983, 1984) and his Frequency Code model, there is a close relationship among the physical size of most animals (large vs small), the vocal pitch height of their vocalizations (low vs. high sounds) and the level of dominance conveyed. The Frequency Code is considered to be biologically determined, but animals, including humans, exploit it for various purposes. For example, Morton (1977) showed how some birds and mammals use low frequency sounds in hostile situations, in order to achieve physical dominance, and high frequency sounds in more pleasant situations, in order to give the impression of being small and vulnerable. Ohala proposed that these patterns can be explained by the assumption that human vocal communication exploits the Frequency Code, for which there is "a cross-species association of high pitch vocalizations with smallness (of the vocalizer), lack of threat, and of low pitch vocalizations with the vocalizer's largeness and threatening intent" (Ohala 1983: 1).

The Frequency Code was later adapted by Gussenhoven (Gussenhoven 2004; Gussenhoven & Jacobs 1998). Accordingly, lower pitched voices are universally encoded to portray "dominant or aggressive individuals" (Gussenhoven 2004: 80) whilst higher pitched voices encode friendliness and even vulnerability. Moreover, according to Gussenhoven, the Effort Code reflects variation in effort, for example a wider excursion of pitch movement is indicative of greater effort. "Speakers exploit this fact by using pitch-span variation to signal meanings that can be derived from the expenditure of effort" (Scherer 1974: 85). It has been suggested that affective meaning of a wide span may be similar to the meaning signaled by low pitch in the Frequency Code, "that of authority due to the metaphor of large size" (Scherer 1974: 88). However, arguably, more effort could also be interpreted as a *lack* of

authority, i.e. having to expend more effort in order to achieve a similar result as someone expending less effort. Gussenhoven also suggested that a wider pitch span might convey more helpfulness on the part of the speaker who uses the wider pitch span. Therefore, it is possible that a higher pitch level, coupled with a wider pitch span, conveys both friendliness, as well as helpfulness, and that these meanings would be reflected in both German and English speakers, given their cross-species meaning (Ohala 1983, 1984).

In relation to the current research, it is especially relevant that a low pitch level may be perceived to signal dominance and aggression, and a narrow pitch span lack of helpfulness, which could arguably be the opposite of how German-Canadians would want to be interpreted by their compatriots in Canada due to the *already* negative impressions associated with the Nazi-German history. As a result, this could mean that German-English bilinguals in the German heritage community in Canada would manipulate their pitch level to convey friendliness through a higher pitch level. This might be more relevant to German males, who would have to lower their pitch level in line with English norms, whilst the German females would *already* target the higher pitch level of English, and hence not run the risk of signaling dominance and aggression, as discussed in more detail below.

5. Pitch in German and English

As already mentioned, pitch can also take on language specific characteristics which can be reflected in attrition and acquisition of prosody (Grazia Busà & Urbani 2011; Mennen & de Leeuw 2014; Mennen et al. 2007, 2012; Ohara 1999; Scherer 1974; Ullakanoja 2007; Van Bezooijen 1995; Willems 1982). Differences in pitch between German and English have been reported in popular media; notably, the female German voice-over for an American sitcom was reported to have a noticeably lower pitch level than the original actress (Eckert & Laver 1994). In further descriptive analyses, it has been documented that "pitch modulation in German is in general much less than in English and many other languages in otherwise comparable situations, which may lead to misjudgements of intention or attitude" (Gibbon 1998: 89). Differences in pitch, which suggest that English females are at one end of a pitch range continuum (highest pitch level and widest pitch span) and German males at the other end of this continuum (lowest pitch level and narrowest pitch span) reflect not only the language specificity of pitch range in German and English, but also gender differences with regard to how pitch range is realized by females and males in German and English. With regard to a quantitative impressionistic analysis of German and English male voices, it was found that German males had significantly "higher pitched voices" (i.e. higher pitch level) than American males, although

no clear difference was found with regard to pitch span for the male German and English speakers (Scherer 1974: 290). For both German and English males, pitch level was found to be strongly associated with pitch span, i.e. a rise in pitch level was associated with a widening of pitch span, although this was "more strongly interrelated" for the German males than for the English males (Scherer 1974: 291).

Recent instrumental research has contrasted pitch range in German and English (Mennen, Schaeffler & Dickie 2014; Mennen et al. 2012; de Leeuw 2010; Scharff-Rethfeldt 2000; Scharff-Rethfeldt, Miller & Mennen 2008). The focus of these studies has largely been on female speech, in part to maintain a unified analysis (i.e. if the males performed contrary to the females, this could have nullified any potential effect) and because the differences between German and English pitch range in impressionistic analyses were more consistent in females than in males: Eckert and Laver (1994) and Gibbon (1998) agreed that German females had a lower pitch level and a narrower pitch span than English females whereas there was no such consensus for males.

In reviewing these studies, an initial investigation looked at speaking fundamental frequency (i.e. pitch level) of female monolingual speakers of British English and German who were between 20 and 40 years of age. The monolingual English speakers were from Newcastle upon-Tyne, whereas those in Germany were from the Stuttgart area (Scharff-Rethfeldt 2000). These monolingual speakers functioned as two respective control groups and were compared with an experimental group of German-English bilinguals who were native German speakers who had been living in the U.K. on a long-term basis. The study revealed that the German monolingual females had a significantly lower speaking F0 than the English monolinguals. Moreover, it was reported that the German native speakers displayed less frequency variation than their English counterparts which suggests a narrower pitch span on the part of the former. In line with the monolinguals' results, the bilinguals used a significantly higher pitch level in English than they did in German. Similarly, it has been reported that there is a tendency for a higher pitch level in English female speech than in German female speech (Mennen et al. 2012, 2014). Moreover, with regard to pitch level, it has been reported that more interpersonal variation occurs in English female speech than in German female speech (Mennen 2007).

If the above studies are summarized, the results suggest that German females have a lower pitch level and a narrower pitch span than English females (Mennen et al. 2012, 2014; Mennen, Schaeffler & Docherty 2007; Scharff-Rethfeldt et al. 2008). In contrast, the findings for men are less clear, but German males potentially have a higher pitch level and a wider pitch span than English males (Scherer 1974; but see Gibbon 1998). Figure 1 attempts to summarize previous research on pitch range in female and male speakers of German and English. With regard to pitch level, the model indicates that in general pitch is higher in English female speech

than in German female speech, whilst pitch is lower in English male speech than in German male speech. With regard to pitch span, the model indicates that pitch span is wider in English female than in German female speech, and wider in German male than in English male speech.
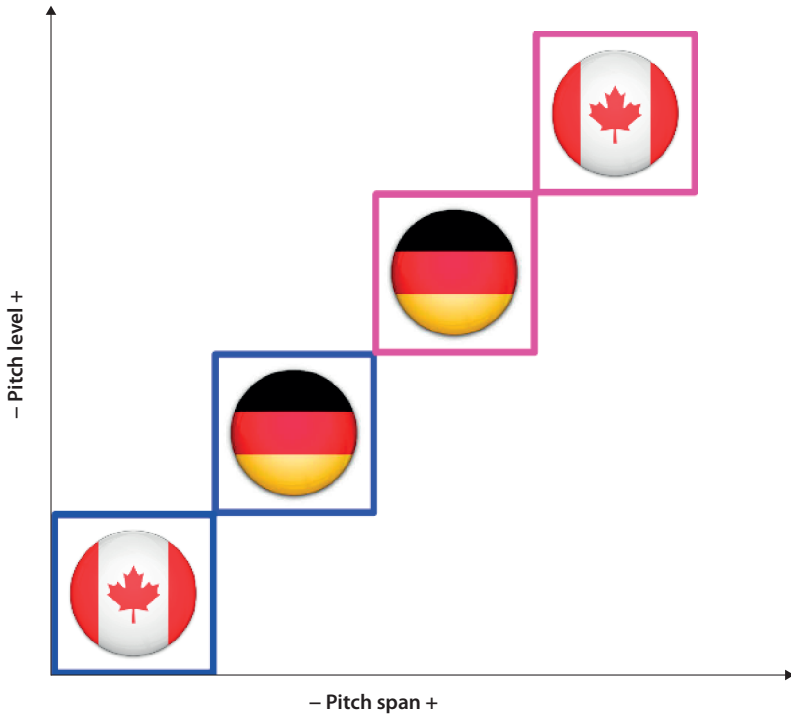
Figure 1. Model of expected pitch range in female and male German and English monolingual speech (pink=female speech; blue=male speech)

What remains unclear is whether and how the German and English females and males react differently to the different language norms of their English L2, and whether this will affect their German native language differently. In particular, the social constellation of the German heritage language community in Vancouver, Canada, in comparison to other immigrant groups in Canada, may have an effect on pitch realization in the bilinguals.

6. Methodology

6.1 Participants

Thirty participants were examined: 10 German L1-English L2 late sequential bilinguals; 10 German monolinguals; and 10 English monolinguals. The age of arrival (AoA) of the bilinguals ranged between 16 to 32 years of age and length of residence (LoR) between 18 and 55 years (see Table 1). AoA was considered to be the same as age of acquisition, as all bilinguals reported that their English was non-existent to rudimentary before their move to Canada in young adulthood, not untypical of school education in Germany at the time.

Table 1. AoA, LoR and gender of bilinguals

Participant	1	2	3	4	5	6	7	8	9	10
AoA	16	19	24	21	32	29	20	32	23	23
LOR	48	22	55	53	29	18	49	48	38	40
Gender	M	F	M	M	F	F	F	F	F	F

The monolinguals were matched with the bilinguals for age at recording (AaR), education, gender, as well as regional accent (see Table 2). The bilinguals and monolinguals self-assessed their German as being in no way or only slightly influenced by a regional accent. The reasoning behind asking for the participants' regional accent was to ensure that the dialectal background of the participants did not confound with attrition of German. Accordingly, one of the original individuals recorded was excluded from the analysis for precisely this reason, as he came from the dialectal region of *Ostfränkisch*.

Additional to the self-assessment, a German phonetician, Edith Braun, who specializes in German dialects, also listened to controlled portions of the participants' recordings and gave her opinion on the extent of regional accent in each recording. Due to time constraints, she did not listen to the recordings in their entirety. This assessment was performed in her home and she wore a headset. It was explained that it was important she listened for the specific regional accent of each participant, even if some of the individuals may have also had English accented speech. She was not told which accent background they might have. Her comments were not decisive in interpreting and categorizing potentially regionally accented speech, but they reinforced the participants' own assessment of their regional accent, as well as my own (namely in both cases that the participants' German was either not influenced by a regional accent, or only weakly influenced).

To summarize her comments, only two of the participants had a slight regional accent. She also listened to the participant with an *Ostfränkisch* dialect and noted his strong regional accent. Participant 1 had, in her opinion, a very slight East or West Prussian accent. In fact, he was born in West Prussia. She also noted that participant 7 had an accent from Southern Germany, potentially coming from either Swabia or the area of *Ostfranken*. In fact, participant 7 grew up in Swabia but lived in Switzerland from 14–34 years of age. For the remaining participants, the phonetician emphasized that no regional accent was prevalent, although she did comment on differing degrees of English accented German speech. Regarding English accented speech, she noted that she could hear from these selected short recordings a strong English accent in participants 4 and 10, and slight English accents in participants 3, 7 and 9. Accordingly, the assumption is that regional accent did not play a strong role in influencing the bilinguals' speech, but note again that accent background was matched with the monolinguals.

Table 2. The age at recording (AaR) of the bilingual experimental group and the control groups (in years)

Bilingual subject	1	2	3	4	5	6	7	8	9	10	Mean
AaR	*72*	*41*	*79*	*73*	*61*	*47*	*69*	*80*	*61*	*63*	*64.6*
German monolingual	11	12	13	14	15	16	17	18	19	20	
AaR	*74*	*37*	*76*	*67*	*57*	*38*	*82*	*71*	*64*	*58*	*62.4*
English monolingual	21	22	23	24	25	26	27	28	29	30	
AaR	*68*	*41*	*72*	*71*	*67*	*55*	*77*	*81*	*67*	*65*	*66.4*

6.2 Data collection

The 'Dog and Duck' story was chosen for the pitch range task, which was translated into German (Mennen et al. 2012, 2007) from the original English version used by Brown and Docherty (1995). This story was considered to be a useful measure of pitch range because an equal amount of direct and indirect speech occurs in both texts and it is characterized by a large amount of voicing, necessary to extract pitch (Mennen et al. 2012, 2007). Subjects were initially shown the one-page story in the respective language and the actual recording commenced thereafter. Reading the story out loud took approximately two minutes and in addition to the preparations beforehand, the recording took approximately seven minutes in each language, conducted in separate blocks with guidance from a native speaker in each language (see de Leeuw 2009 for more information).

6.3   Measuring pitch range

Pitch level was measured in Praat (Boersma & Weenink 2010), using the analysis settings as recommended in the Praat manual. Accordingly, for women pitch floor was set to 100 Hz while the pitch ceiling was set to 500 Hz. For men, pitch floor was set to 75 Hz while the ceiling was set to 300 Hz. For pitch span, the difference between the 90th and 10th percentile range (80% Range) in semitones (ST), inter-quartile range (IQR) (i.e. the limits within which the middle 50% of an ordered set of observations fall in ST), and +/− 2 standard deviations around the mean (SD4) in ST, were obtained. For pitch level, mean and median F0 (Hz) were measured. These measurements were chosen as they had already been proven successful in similar German to English comparisons of pitch range (Mennen et al. 2012, 2007). Based on these settings, mean and median F0 in Hz, 80% Range in ST, IQR in ST, and SD4 in ST were measured in the speech samples. The first value represented pitch level, whereas the latter were various quantifications of pitch span. As the results from the different pitch level measurements were similar, as well as the results from the different pitch span measurements, only mean F0 (Hz) is reported for pitch level, and 80% Range for pitch span, which have similarly been reported in previous research (Mennen et al. 2014, 2012; Ordin & Mennen 2017).

6.4   Hypotheses

Based on our knowledge of monolingual pitch range in female and male German speakers, the following hypotheses were made:

a. Bilingual females will have a higher pitch level in German than monolingual German females due to the acquisition of a higher pitch level in English;
b. Bilingual males will have a lower pitch level in German than monolingual German males due to the acquisition of a lower pitch level in English;
c. Bilingual females will have a wider pitch span in German than monolingual German females due to the acquisition of a wider pitch span in English;
d. Bilingual males will have a narrower pitch span in German than monolingual German males due to the acquisition of a narrower pitch span in English.

In addition, it was also considered relevant to bear in mind that the German-English bilinguals would intend to signal dissociation from Nazi-German history (Lieb 2008), and that pitch may have been implemented as a linguistic tool to do so.

7.   Results

7.1   Pitch level

Testing whether there was a significant difference between the pitch level of the bilinguals and monolinguals was conducted separately for the male and female participants because opposite predictions were made for males and females. Non-parametric tests were chosen for these analyses because the assumption of homogeneity of variance, as indicated by Levene's test, was violated.

The results suggested that there was a tendency for the pitch level of the German monolingual females to be lower than that of the English monolingual females: the averaged mean F0 of the German females was approximately 185 Hz, whereas the averaged mean F0 of the English females was just over 198 Hz (see Figure 2). In comparison with the monolinguals, there appeared to be a tendency for the female bilinguals to pattern intermediately with regard to mean F0 (average of 193 Hz in German and 194 Hz in English). However, the Kruskal-Wallis tests for mean F0 in females was not significant ($H(2) = 1.09$, $p = .58$), which may have been due to high interspeaker variation in English, or to small participant numbers. It is noticeable that there was less variation with regard to pitch level in the female German monolinguals than in the female English monolinguals. The standard deviation of mean F0 for the female monolinguals in German was just over 18 Hz, whereas in English it was just over 27 Hz. Moreover, there was more variation in the bilingual female pitch level in both English and German than in the German female pitch level.
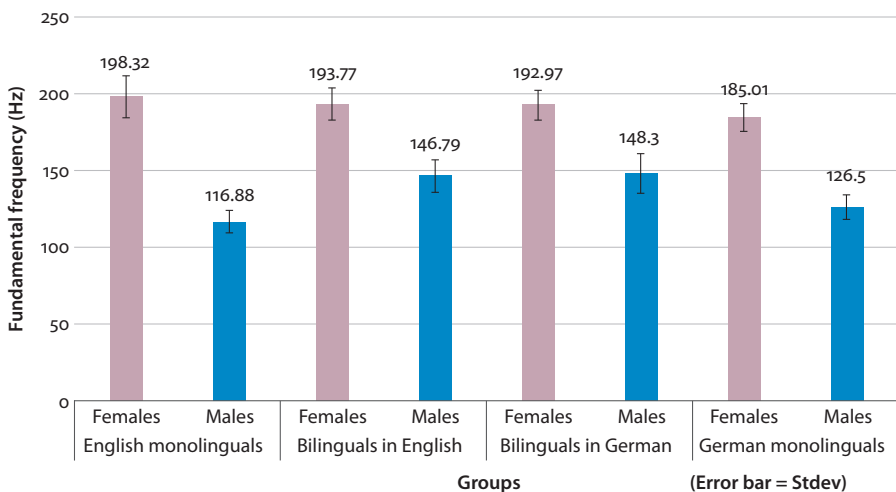
Figure 2. Mean pitch level in monolinguals and bilinguals

As expected, German monolingual males had on average a higher pitch level than English monolingual males (see Figure 2). More specifically, German males had an average mean F0 of almost 127 Hz, whereas the averaged English male mean F0 was just under 117 Hz. Surprisingly, the male bilingual mean F0 was higher (average of 148 Hz in German and 147 Hz in English) than that of *both* male monolingual groups. In contrast to the analyses of the females, the Kruskal-Wallis tests performed on the pitch level of the male speakers were closer to reaching significance (mean F0, $H(2) = 5.07$, $p = .09$), most likely driven by the high pitch level of the male bilinguals in German and English.

These findings suggest that had the group sizes been larger, significant differences may have emerged in line with the language specific trends in the pitch level of the monolinguals and bilinguals. Nonetheless the finding that pitch level was higher in the German and English of bilingual males countered the hypotheses, and will be explored in the discussion.

7.2   Pitch span

As expected, the pitch span of the German monolingual females was on average narrower than the pitch span of the English monolingual females (see Figure 3). In the former group, for example, an 80% Range of 8.07 ST was observed, whereas in the latter group, this was 9.78 ST. However, this did not prove to be significant: for 80% Range, $H(2) = .45$, $p = .80$, which was again potentially due to the small group sizes, as well as to the high amount of variation within the female English monolingual group. As observed in the analysis of pitch level, the pitch span of the
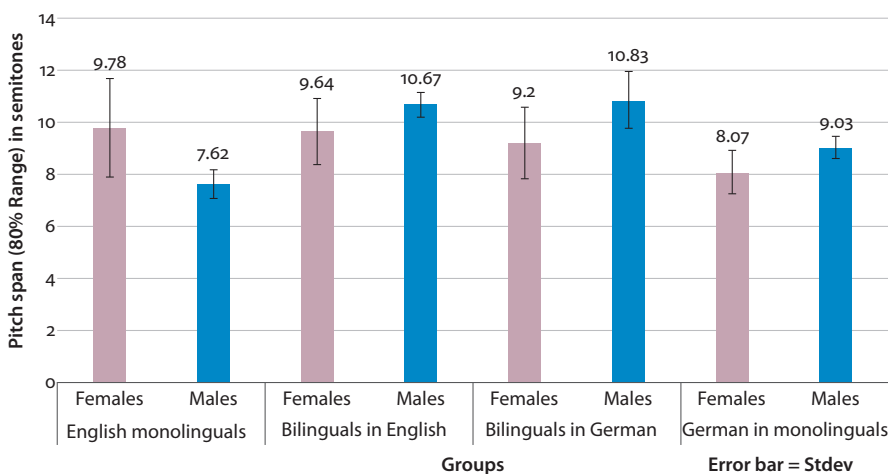
Figure 3. Pitch span (80% Range) in monolinguals and bilinguals

bilingual females was on average intermediate to that of the monolingual females (see Figure 3).

As expected, a wider pitch span on the part of the German males was observed than on the part of English males (see Figure 3). In fact, the German male controls had a wider pitch span than the German female controls. In contrast, the English female monolinguals displayed a rather wider pitch span than the English male controls in all pitch span measurements. The difference between pitch span was also less between the German monolingual females and males than between the English monolingual females and males.

The bilingual males had an even wider pitch span than did the German monolingual males. In fact, the bilingual male pitch span was wider in both of their languages than the bilingual females: 80% Range in the German of the bilingual males was 10.83 ST whereas in English it was 10.67 ST. In contrast to the women, the Kruskal-Wallis tests for the men were significant (for the dependent variable of 80% Range, $H(2) = 5.85$, $p < .05$). These results were followed up by conducting two Mann-Whitney tests. Test 1 investigated differences in 80% Range between the monolingual groups and Test 2 investigated differences in 80% Range between the German of the bilingual and the German monolingual group. Test 1 was significant ($U = 0.0$, $p = .046$), but Test 2 was not significant ($U = 2.5$, $p = .37$). Given that the male groups were quite small, it is suggested that with a larger sample size there may have indeed been a significant difference between 80% Range in the monolingual males and bilinguals. However, it is evident here that the German male pitch span was wider than the English male pitch span – and yet nonetheless, there was a trend for the German of the bilinguals to have an even wider pitch span. This finding will be explored in the discussion.

7.3   Bilingual variation in pitch range

In Figure 4, mean F0 is displayed on the x-axis and 80% Range on the y-axis for female and male bilinguals and monolinguals.

Firstly, it can be seen from the scatterplot (Figure 4) that the English monolingual females were spread out over a slightly wider area of pitch level along the fundamental frequency x-axis than the German females. With regard to pitch span along the 80% Range on the y-axis, the German and English female monolinguals patterned similarly. However, the bilingual females widened their pitch span when speaking English, in contrast to when speaking German, in most cases, whilst there was not as much movement with regard to their pitch level. One bilingual female had a wider pitch span in German than in English, which may have contributed to the overall lack of significance in females due to averaging effects.
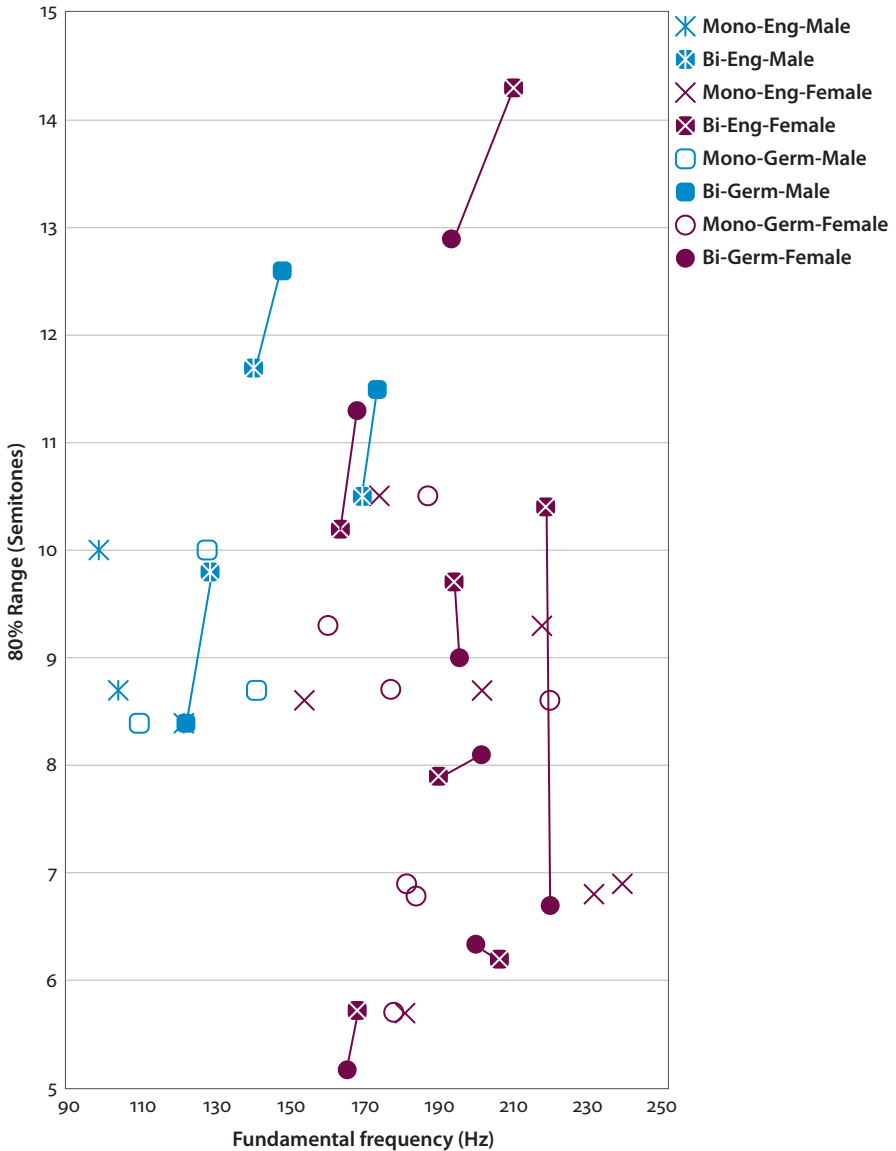
Figure 4. Scatterplot of pitch span (80% Range) over pitch level (F0) for monolinguals and bilinguals in German and English

With regard to the men, it can be observed that the German and English monolingual males patterned relatively similarly, although the German men tended to have a higher pitch level on the x-axis than the English men, and their pitch span on the y-axis was comparable with the English men. What is noticeable is that two

of the three bilingual males widened their pitch span and increased their pitch level in their German, in comparison to their English. One of these men even patterned with regard to both pitch range dimensions within the female area. For both of these two male bilinguals, their pitch span increased in German in comparison to English, which was the opposite of the effect observed in most of the female bilinguals, who evidenced a wider pitch span in English than in German, save one exception. Additionally, there was a positive relationship between pitch level and span for men. In fact, a Pearson's correlation test indicated a highly significant relationship between mean F0 and 80% Range ($r = .76$, p (one-tailed) < .01): the higher the pitch level was in men, the wider their pitch span.

To summarize the bilingual variation in pitch range, it is possible to deduce that, in general, the bilingual females tended to have a wider pitch span in their English than in their German, but that in the bilingual males the opposite was true: They tended to have a wider pitch span in their German than in their English. Moreover, in general, there was not much movement with regard to pitch level in German and English for the bilingual females, but the men tended to have a higher pitch level in German than in English. For both of these men, however, pitch level was noticeably higher in both their German and English than in either of the monolingual groups, and pitch span was noticeably wider in both their German and English than in either of the monolingual groups.

8.   Discussion

With regard to the monolinguals, the results suggested that German females tended to use a lower pitch level than the English females, but that the German males tended to use a higher pitch level than the English males. This was in line with the model in Figure 1. With regard to pitch span in the monolinguals, the German females tended to use a narrower pitch span than the English females, but the German males tended to use a wider pitch span than the English males. Again, these results were consistent with the model in Figure 1 and are largely in line with previous research which has examined German and English bilingual speech (Mennen et al. 2014, 2007; Scharff-Rethfeldt et al. 2008; Scherer 1974). In sum, the results from the monolinguals of the present study confirm that females and males perform more closely to one another in terms of pitch range in German than do females and males in English. Given similar anatomy of German and English males, and similar anatomy of German and English females, these differences were considered to have arisen due to different social and cultural expectations, reflected in German and English.

In terms of the hypotheses, the bilingual German-English females tended to have a higher pitch level in German than monolingual German females which may have been due to the acquisition of the higher pitch level in English, although this was not confirmed statistically. Moreover, the German-English bilingual females tended to have a wider pitch span in German than monolingual German females, again potentially due to the acquisition of a wider pitch span in their English, although this could also not be substantiated statistically. In line with the Frequency Code, a higher pitch level and a wider pitch span could alleviate any efforts on the part of the females in the German heritage community to sound friendly, helpful, and unaggressive, therefore dissociating themselves from any potential aggressive images of Germans in Canada (Lieb 2008). Essentially, as the bilingual females successfully adopted the higher pitch level and wider pitch span, characteristic of English female pitch, they would not have to "worry" about sounding aggressive, because their adoption of the higher pitch level and wider pitch span of female English speech would already convey friendliness and helpfulness.

The German-English bilingual males evidenced a surprising result, contrary to the hypotheses, but potentially in line with the Frequency Code. Instead of having a lower pitch level in German than monolingual German males, as expected due to the acquisition of a lower pitch level in English, the exact opposite was observed: two of the three bilingual males in Canada evidenced a salient higher pitch level than that of either the monolingual male groups. Similarly, rather than the bilingual males in Canada having a narrower pitch span in German than monolingual German males, due to the acquisition of a narrower pitch span in English, as hypothesized, they had a much wider pitch span in both their German and English than either of the monolingual male groups, and one bilingual male even performed within the female pitch range norms.

This finding can potentially be interpreted in relation to the social and political situation of German immigrants to Canada. If a low pitch level is associated with dominance and aggression (Gussenhoven 2004; Gussenhoven & Jacobs 1998; Ohala 1983, 1984), as the Frequency Code suggests, the German males in Canada could have attempted to dissociate themselves from these characteristics by increasing their pitch level to convey to Canadian interlocutors that they were *not* dominant and aggressive, the traits Canadians would commonly associate with Nazi-Germany. Instead, through implementing a higher pitch level, the German males would convey friendliness, and potentially helpfulness, through the implementation of a wider pitch span, and thereby dissociate themselves from Nazi-Germany. These pitch changes would be implemented in their English, but then be taken up in their German native language as well, in the form of attrition. The desire to convey friendliness and helpfulness would potentially be more

pressing for males than for females because the German-English females tended to increase their pitch level and widen their pitch span anyway. Through a higher pitch level, the bilingual males in Canada would be able to signal dissociation from an "immediate connection with the Hitler dictatorship" (Lieb 2008: 310) and instead convey gentler attributes. Through the higher pitch level and wider pitch span, the German-English bilingual men would successfully play down traits associated with "Germanness" in Canada, which have led to disadvantages, and even abuse, and created an environment in which Germans have attempted to hide their German heritage (Lieb 2008).

As such, it may be useful to draw on the social constellation of the German heritage language community in Vancouver, BC, which has largely kept a low profile. In doing so, it requires that different societal groups are examined (e.g. here females and males, those of German heritage, and those not) to examine how various linguistic variables are indexed differently within the same community. Such methodologies are usually implemented within monolingual sociolinguistic analyses (e.g. Cheshire, Kerswill, Fox & Torgersen 2011; Eckert & Rickford 2001; Labov 1963; Tagliamonte & D'Arcy 2009; Trudgill 1986) but there is potential for similar strategies to be applied in bilingual research, as has been attempted here. New research examining the Japanese-English community in London is furthering this goal, with preliminary findings suggesting that Japanese-English bilingual females dissociate themselves from stereotypical forms of Japanese femininity by lowering their pitch level in Japanese and English (Passoni, Mehrabi, Levon & de Leeuw 2018). Furthermore, some studies have found that in men, either gay themselves, or involved in gay activism, pitch varies due to both the conversational setting, as well as the topic at hand within monolinguals, as a function of social meaning (Levon 2009; Podesva 2007), which is also in line with the present results. It seems plausible that men in the German heritage community might also manipulate pitch to convey social, and potentially political, meaning. If human vocal communication exploits the Frequency Code across language, an increase in pitch level and a widening of span would immediately indicate friendliness and helpfulness, which German males in the German heritage language community might want to signal, in order to dissociate from any potential connection with the Hitler dictatorship (Lieb 2008).

In large immigrant communities, such as the German-English heritage language community of Vancouver, Canada, in which the heritage language is spoken at home and in the community, bilinguals who have undergone attrition are inextricably linked to the larger community as such (Montrul 2008; Polinsky 2008). It may be that, as time passes since the atrocities of WWII, Germans abroad become less associated with the "enemy", such that, ultimately, the proposed effects evidenced in this study diminish in younger generations. However, if some of the original

properties of German pitch are not available to the heritage language learner, i.e. the children of the community as a result of attrition in the parent generation (Rothman 2007), heritage language children would arguably never be able to acquire that which was not available to them, unless there was contact with the source language, i.e. through family and friends still in the country of origin. If it becomes less relevant in Canada for younger generations to signal dissociation from the atrocities of WWII, it may be possible to find new norms arising within the German heritage language community over time. Future research would need to assess whether the trends observed here hold for larger samples, across different generations, and aim to understand the social significance of pitch patterns in this community better by collecting additional ethnographic data on bilinguals' attitudes and gender identities (see, e.g., Passoni et al. 2018).

However, alternative explanations for the observed results are possible as well. For example, it may be that the bilingual males predominantly interacted with females in their environment, and hence aimed to emulate female speech patterns, including a higher pitch level and wider pitch spans. This interpretation seems relatively unlikely though, as it would nevertheless be the case that the bilingual males would interact with men as well, and the question therefore still arises which is *why* the bilingual males would emulate female speech patterns, rather than male speech patterns. Again, the Frequency Code may offer an explanation. *Both* the females and males may have targeted the higher pitch level and wider pitch span in order to convey friendliness and helpfulness, to quickly dissociate with any negative image related to their Germanness in Canada. Nevertheless, a consequence of the wider pitch span and higher pitch level in the German bilingual males would be to potentially be perceived as less masculine by Canadian men. Perhaps, given the reported disadvantages and sometimes abuse experienced by Germans in Canada (Lieb 2008), it would still be more advantageous to convey friendliness and helpfulness as a German male, rather than producing a lower pitch and narrower pitch span.

In summary, these findings are inconclusive, yet they raise interesting questions with regard to gender specific differences in pitch range in German-English bilinguals of the German heritage language community in Vancouver, Canada. The findings suggest that although L2 acquisition is influential in attrition processes, the societal and political configuration of the heritage language community also plays a role in determining the speech patterns of the members of that community.

Acknowledgements

References

Au, T. K., Knightly, L. M., Jun, S.-A. & Oh, J. S. 2002. Overhearing a language during childhood. *Psychological Science* 13(3): 238–243. https://doi.org/10.1111/1467-9280.00444

Boersma, P. & Weenink, D. 2010. *PRAAT. Doing Phonetics by Computer* [Computer Program]. University of Amsterdam. <http://www.praat.org>

Bowers, J. S., Mattys, S. L. & Gage, S. H. 2009. Preserved implicit knowledge of a forgotten childhood language. *Psychological Science* 20(9): 1064–1069. https://doi.org/10.1111/j.1467-9280.2009.02407.x

Brown, A. & Docherty, G. J. 1995. Phonetic variation in dysarthric speech as a function of sampling task. *International Journal of Language & Communication Disorders* 30(1): 17–35. https://doi.org/10.3109/13682829509031320

Cherciov, M. 2013. Investigating the impact of attitude on first language attrition and second language acquisition from a Dynamic Systems Theory perspective. *International Journal of Bilingualism* 17(6): 716–733. https://doi.org/10.1177/1367006912454622

Cheshire, J., Kerswill, P., Fox, S. & Torgersen, E. 2011. Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics* 15(2): 151–196. https://doi.org/10.1111/j.1467-9841.2011.00478.x

de Leeuw, E. 2009. When Your Native Language Sounds Foreign: A Phonetic Investigation into First Language Attrition. PhD dissertation, University of Edinburgh.

de Leeuw, E. 2010. Measuring language-specific phonetic settings. *Second Language Research* 26(1): 13–41. https://doi.org/10.1177/0267658309337617

de Leeuw, E. 2019. Phonetic attrition. In *The Oxford Handbook of Attrition*, M.S. Schmid et al. (eds), 204–217. Oxford: OUP.

Dominguez, L. & Hicks, G. 2016. Synchronic change in a multidialectal Spanish community: Evidence from null and postverbal subjects. In *Inquiries in Hispanic Linguistics: From Theory to Empirical Evidence* [Issues in Hispanic and Lusophone Linguistics 12], A. Cuza, L. Czerwionka & D. Olson (eds), 53–72. Amsterdam: John Benjamins. https://doi.org/10.1075/ihll.12.04dom

Eckert, H. & Laver, J. 1994. *Menschen und ihre Stimmen*. Weinheim: Psychologie Verlags Union.

Eckert, P. & Rickford, J. R. 2001. *Style and Sociolinguistic Variation*. Cambridge: CUP.

Gibbon, D. 1998. Intonation in German. In *Intonation Systems. A Survey of Twenty Languages*, D. Hirst & A. Di Cristo (eds), 78–95. Cambridge: CUP.

Grazia Busà, M. & Urbani, M. 2011. A cross-linguistic analysis of pitch range in English L1 and L2. In *Proceedings of the 17th International Conference of Phonetic Sciences*, 380–383. Hong Kong: City University of Hong Kong.

Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge: CUP.

https://doi.org/10.1017/CBO9780511616983

Gussenhoven, C. & Jacobs, H. 1998. *Understanding Phonology*. London: Arnold.

Hayward, K. 2000. *Experimental Phonetics: An Introduction*. Harlow: Pearson Education.

Hewlett, N. & Beck, J. 2006. *An Introduction to the Science of Phonetics*. Mahwah NJ: Lawrence Erlbaum Associates.

Kagan, O. & Dillon, K. 2001. A new perspective on teaching Russian: Focus on the heritage learner. *The Slavic and East European Journal* 45(3): 507–518. https://doi.org/10.2307/3086367

Kupisch, T. & Rothman, J. 2018. Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism* 22(5): 564–582. https://doi.org/10.1177/1367006916654355

Labov, W. 1963. The social motivation of a sound change. *Word* 19(3): 273–309. https://doi.org/10.1080/00437956.1963.11659799

Ladd, D. R. 2008. *Intonational Phonology*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511808814

Laver, J. 1980. *The Phonetic Description of Voice Quality*. Cambridge: CUP.

Levon, E. 2009. Dimensions of style: Context, politics and motivation in gay Israeli speech. *Journal of Sociolinguistics* 13(1): 29–58. https://doi.org/10.1111/j.1467-9841.2008.00396.x

Lieb, C. 2008. German diaspora experiences in British Columbia after 1945. In *German Diasporic Experiences: Identity, Migration, and Loss*, M. Schulze, J. M. Skidmore, D. G. John, G. Liebscher & S. Siebel-Achenbach (eds), 305–316. Waterloo: Wilfrid Laurier University Press.

Linville, S. E. 1996. The sound of senescence. *Journal of Voice* 10(2): 190–200. https://doi.org/10.1016/S0892-1997(96)80046-4

Mennen, I. 2007. Phonological and phonetic influences in non-native intonation. In *Non-native Prosody. Phonetic Description and Teaching Practice*, J. Trouvain & U. Gut (eds), 53–76. Berlin: Mouton de Gruyter.

Mennen, I. & de Leeuw, E. 2014. Beyond segments. *Studies in Second Language Acquisition* 36(2): 183–194. https://doi.org/10.1017/S0272263114000138

Mennen, I., Schaeffler, F. & Dickie, C. 2014. Second language acquisition of pitch range in German learners of English. *Studies in Second Language Acquisition* 36: 303–329. https://doi.org/10.1017/S0272263114000023

Mennen, I., Schaeffler, F. & Docherty, G. 2012. Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America* 131(3): 2249–2260. https://doi.org/10.1121/1.3681950

Mennen, I., Schaeffler, F. & Docherty, G. J. 2007. Pitching it differently: a comparison of the pitch ranges of German and English speakers. In *Proceedings of the 16th International Congress on Phonetic Sciences*, 1769–1772. Saarbrücken: Universität des Saarlandes.

Montrul, S. 2008. *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor* [Studies in Bilingualism 39]. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.39

Morton, E. S. 1977. On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist* 111(981): 855–869. https://doi.org/10.1086/283219

Neppert, J. M. 1999. *Elemente einer akustischen Phonetik*, 4th edn. Hamburg: Buske.

Nishio, M. & Niimi, S. 2008. Changes in speaking fundamental frequency characteristics with aging. *Folia Phoniatrica et Logopaedica* 60(3): 120–127. https://doi.org/10.1159/000118510

Ohala, J. J. 1983. Cross-language use of pitch: An ethological view. *Phonetica* 40(1): 1–18. https://doi.org/10.1159/000261678

Ohala, J. J. 1984. An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41(1): 1–16. https://doi.org/10.1159/000261706

Ohara, Y. 1999. Performing gender through voice pitch: A cross-cultural analysis of Japanese and American English. In *Wahrnehmung und Herstellung von Geschlecht*, U. Pasero & F. Braun (eds), 105–116. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-322-89014-6_8

Opitz, C. 2013. A dynamic perspective on late bilinguals' linguistic development in an L2 environment. *International Journal of Bilingualism* 17(6): 701–715. https://doi.org/10.1177/1367006912454621

Ordin, M. & Mennen, I. 2017. Cross-linguistic differences in bilinguals' fundamental frequency ranges. *Journal of Speech, Language, and Hearing Research* 60(6): 1493–1506. https://doi.org/10.1044/2016_JSLHR-S-16-0315

Pallier, C., Dehaene, S., Poline, J.-B., LeBihan, D., Argenti, A.-M., Dupoux, E. & Mehler, J. 2003. Brain imaging of language plasticity in adopted adults: Can a second language replace the first? *Cerebral Cortex* 13(2): 155–161. https://doi.org/10.1093/cercor/13.2.155

Passoni, E., Mehrabi, A., Levon, E. & de Leeuw, E. 2018. Bilingualism, pitch range and social factors: preliminary results from sequential Japanese-English bilinguals. In *Proceedings of the 9th International Conference on Speech Prosody 2018*, 1–5. Poznań: Adam Mickiewicz University.

Patterson, D. 2000. *A Linguistic Approach to Pitch Range Modelling*. Edinburgh: EUP.

Podesva, R. J. 2007. Phonation type as a stylistic variable: The use of falsetto in constructing a persona1. *Journal of Sociolinguistics* 11(4): 478–504. https://doi.org/10.1111/j.1467-9841.2007.00334.x

Polinsky, M. 2008. Gender under incomplete acquisition: Heritage speakers' lnowledge of noun categorization. *Heritage Language Journal* 6(1): 40–71. <http://dash.harvard.edu/handle/1/3382970> (9 December 2018).

Rothman, J. 2007. Heritage speaker competence differences, language change, and input type: Inflected infinitives in Heritage Brazilian Portuguese. *International Journal of Bilingualism* 11(4): 359–389. https://doi.org/10.1177/1367006907011004201

Rothman, J. & Treffers-Daller, J. 2014. A prolegomenon to the construct of the native speaker: Heritage speaker bilinguals are natives too! *Applied Linguistics* 35(1): 93–98. https://doi.org/10.1093/applin/amt049

Scharff-Rethfeldt, W. 2000. Speaking Fundamental Frequency Differences in the Language of Bilingual Speakers. PhD dissertation, University of Newcastle upon Tyne.

Scharff-Rethfeldt, W., Miller, N. & Mennen, I. 2008. Speaking fundamental frequency differences in highly proficient bilinguals of German/English. *Sprache, Stimme, Gehör* 32(3): 123–128. https://doi.org/10.1055/s-0028-1083799

Scherer, K. R. 1974. Voice quality analysis of American and German speakers. *Journal of Psycholinguistic Research* 3(3): 281–298. https://doi.org/10.1007/BF01069244

Schmid, M. S. 2011. *Language Attrition*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511852046

Schmid, M. S. & Köpke, B. 2007. Bilingualism and attrition. In *Language Attrition: Theoretical Perspectives* [Studies in Bilingualism 33], B. Köpke, M. S. Schmid, M. Keijzer & S. Dostert (eds), 1–7. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.33.02sch

Tagliamonte, S. A. & D'Arcy, A. 2009. Peaks beyond phonology: Adolescence, incrementation, and language change. *Language* 85(1): 58–108. https://doi.org/10.1353/lan.0.0084

Todd, D. 2011. Ethnic mapping 5: Find Metro's Dutch, Germans, Iranians and Italians. *Vancouver Sun 19 October 2011*. <http://vancouversun.com/news/staff-blogs/ethnic-mapping-5-find-metros-dutch-blacks-germans-and-iranians> (9 December 2018).

Trudgill, P. 1986. *Dialects in Contact*. Oxford: Blackwell.

Ullakanoja, R. 2007. Comparison of pitch range in Finnish (L1) and Russian (L2). In *Proceedings of the 16th International Congress on Phonetic Sciences*, 1701–1704. Saarbrücken: Universität des Saarlandes.

Van Bezooijen, R. 1995. Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech* 38(3): 253–265. https://doi.org/10.1177/002383099503800303

Willems, N. 1982. *English Intonation from a Dutch Point of View*. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110853469

Does extensive L2 exposure trigger L1 attrition of perfective and durative aspect marking in Mandarin Chinese?

Shi Zhang
University of Stirling

This study concerns whether extended exposure to a second language would lead to first language (L1) attrition, i.e. changes of L1 linguistic behavior/knowledge. An acceptability judgement task, which examined the perceptive knowledge of perfective and durative aspect marking in Mandarin Chinese, was employed, and the performance of 14 Mandarin-English bilinguals in the UK was examined. The results did not suggest that the bilinguals showed L1 attrition in perceiving perfective/durative aspect marking. The paper also discusses how research on heritage language acquisition could benefit from L1 attrition research.

Keywords: L1 attrition, Chinese, perfective aspect, durative aspect, bilingualism

1. Introduction

In recent years, research on heritage languages (HLs) and heritage speakers (HSs) has increased rapidly. HLs usually refer to the non-societal and non-majority languages in a certain linguistic environment, such as Spanish in the US (Valdés 2005). HSs are often defined as bilinguals who have acquired the HL – as a first language (L1) – and a majority language – as another L1 or a second language (L2) – naturalistically in early childhood (Pascual y Cabo & Rothman 2012; Rothman & Treffers-Daller 2014).

Various studies on HL acquisition have shown how HL acquisition might differ from monolingual L1 acquisition and typical L2 acquisition in terms of routes and outcomes (e.g., Montrul 2008, 2016; Polinsky 2011). A number of these studies argued that HSs might have been exposed to HL "input from the first generation which has already undergone changes under the influence of the (L2) majority language" (Kupisch 2013: 207), and this potentially changed quality of HL input

might have an important impact on the outcomes of HL acquisition. For instance, Rothman (2007) reported that advanced HSs of Brazilian Portuguese with no formal education in Brazilian Portuguese were unlikely to acquire monolingual-like knowledge about inflected infinitives, because this structure was only available in formal, but not colloquial input in Brazilian Portuguese. Another study by Brehmer & Kurbangulova (2016) also observed a connection between the type of input and performance in HSs. They reported that in a number of tasks concerning pronunciation, inflectional morphology and vocabulary, the performance of the young HSs of Russian mirrored that of their parents, who provided the main source of early HL input.

Given the fact that the quality of HL input plays an important role in shaping the outcomes of HL acquisition, and that parents provide the main source of HL/L1 input to HSs during early childhood, it is reasonable to argue that, in order to further understand the relationship between the quality of HL input and the outcomes of HL acquisition, it is necessary to study what changes the L1s of the parents have undergone after being extensively exposed to input in the majority language (L2) and reduced input in their L1s. In other words, studying L1 attrition[1] among adult migrants who have lived in an L2 environment for an extended period will help us to better understand the quality of HL input, which might further affect HL acquisition.

Previous studies have investigated various aspects of L1 attrition, including grammatical knowledge, phonetics, phonology, processing and social attitudes among adult and child speakers of different languages, such as German, Italian, Greek or Turkish.[2] As an HL spoken by millions of people outside Mainland China and Taiwan, overseas speakers of Mandarin Chinese (henceforth Mandarin) have also received increasing interest in research. However, the existing studies concerning overseas Mandarin speakers mostly focused on sociolinguistic and educational topics, such as how overseas Mandarin speakers form their identity, parents' and HSs' attitudes towards Mandarin, and the maintenance of reading/writing skills (e.g., Ding 2013; Li 2006; Luo 2015), and only a very limited number of studies have experimentally studied the L1 attrition of Mandarin grammatical knowledge among adult Mandarin speakers living overseas.[3] Furthermore, while several

---

1. In this paper, we will use the term *L1 attrition* to refer to the change of any aspect of an L1, as long as the change is associated with long-term, extensive exposure to an L2 and/or lack of exposure to the L1. For a discussion about the definition of L1 attrition, see Schmid & Köpke (2017).

2. See Schmid (2016) for an annotated bibliography of these studies.

3. But see Aalberse, Zou & Andringa (2017) on demonstrative pronouns, and Hui (2012) on reflexives.

studies reported that Mandarin HSs deviated from monolinguals in producing aspect markers (Jia & Bayley 2008; Ming & Tao 2008; Shi 2011), there seems to be few studies on L1 attrition in aspect marking in Mandarin. Therefore, it remains unclear whether Mandarin HSs' problems with using aspect markers are associated with the quality of HL input.

Aiming to gain some preliminary insight into the L1 attrition in aspect marking in Mandarin, this small scale pilot study looks into whether adult Mandarin speakers who have lived in the UK for an extended period would show attrition in perfective and durative aspect marking in Mandarin. Previous studies found that young HSs of Mandarin with L2 English exhibited non-monolingual-like performance of producing perfective aspect marking (Jia & Bayley 2008; Ming & Tao 2008; Shi 2011), but they did not concern the quality of the HL input received by those HSs, or the relationship between HL input and the outcomes of HL acquisition. By investigating whether Mandarin speakers in the UK would demonstrate attrition of perfective and durative aspect marking, this study can provide some empirical evidence that will expand our understanding about how the L1 knowledge of aspect marking in Mandarin might be affected in an L2 environment, and help us to predict Mandarin HSs' acquisition of aspect marking in the future.

This paper is organized as follows: The second section will introduce the aspect system of Mandarin Chinese, and review the existing studies related to the attrition of aspect marking by overseas Mandarin speakers. The third section will present the research question and hypothesis of this study, and the fourth section will reflect on the methodology of this study. The last three sections will provide the results, discussions and a conclusion.

2. Background

2.1 The aspect system of Mandarin

In Mandarin, aspect marking is an important way of expressing temporal information. Research on aspect in Mandarin usually adopts a two-component approach to aspect, and assumes a distinction between lexical aspect and grammatical aspect (Comrie 1976; Smith 1997; see also Vendler 1967; Xiao & McEnery 2004). Lexical aspect is inherent in verbs and predicates, and concerns the internal temporal features of situations. For instance, in English, *run a mile* encodes a natural end point, so it has an internal feature of [+Telic] in terms of lexical aspect; by contrast, *run* does not encode such an end point and thus has an internal feature of [−Telic]. In this study, we adopt Xiao & McEnery's (2004) classification of lexical aspect in Mandarin (see Table 1) because it is a working one for the purpose of this study,

which focuses on the interaction between lexical and grammatical aspect. However, we are aware that there is an ongoing debate about how to classify lexical aspect in Mandarin (compare, e.g., Klein, Li, Hendriks & Language 2000; Peck, Lin & Sun 2013; Smith 1997).

Table 1. Xiao & McEnery's (2004) system of lexical aspect in Mandarin Chinese

Classes	[±Dynamic]	[±Durative]	[±Bounded]	[±Telic]	[±Result]	Example
Activity (ACT)	+	+	−	−	−	*run*
Semelfactive (SEM)	+	−	±	−	−	*cough*
Accomplishment (ACC)	+	+	+	+	−	*write*
Achievement (ACH)	+	−	+	+	+	*arrive*
Individual-level state (ILS)	−	+	−	−	−	*resemble*
Stage-level state (SLS)	±	+	−	−	−	*be busy*

Grammatical aspect concerns how a temporal situation is viewed from a speaker's perspective, and is realized differently in languages. In Mandarin, grammatical aspect is realized by aspect markers, and researchers widely acknowledge four aspect markers: the perfective marker *le*, the experiential marker *guo*, the progressive marker *zai*, and the durative marker *zhe* (see Klein et al. 2000; Liu 2015; Wiedenhof 2015; Xiao & McEnery 2004). Table 2 presents the meaning of these four aspect markers:

Table 2. Grammatical aspect markers in Mandarin Chinese (adapted from Duff & Li 2002)

Class	Markers	Meaning	Examples
Perfective	*le*	bounded, perfective	*Ta qu le Shanghai.* "He went to Shanghai."
	guo	experiential	*Ta qu guo Shanghai.* "He has been to Shanghai."
Imperfective	*zai*	progressive	*Ta zai chi wufan.* "He is having lunch."
	zhe	stative, durative, progressive situation	*Ta chang zhe ge xizao.* "He sang while taking a shower."

As shown in the examples in Table 2, lexical aspect encoded in verbs/predicates and grammatical aspect encoded in aspect markers interact to deliver temporal information in Mandarin (compare *qu le Shanghai* with *qu guo Shanghai* in Table 2). Existing research has found that some aspect markers tended to co-occur more frequently with verbs/predicates of certain lexical aspects, while not co-occurring with verbs of other lexical aspects. For example, the durative marker *zhe* indicates that a situation is viewed as enduring or continuing, so it tends to co-occur with

verbs/predicates with a [+Durative] feature, such as ACTs, ACCs, ILSs[4] and SLSs, rather than those with a [−Durative] feature, such as SEMs or ACHs. By contrast, when appearing at verb-final positions, the perfective marker *le* indicates the completion or termination of a situation with reference to time in the past, present or future, and it tends to co-occur with ACHs, ACCs, ACTs, but not SEMs, ILSs, or SLSs (Xiao & McEnery 2004). In this study, we will focus on potential L1 attrition in the interaction between different lexical aspects and two types of aspect markers, namely the perfective marker *le* and the durative marker *zhe*. It should be noted that this study only concerns the verb-final *le*, as sentence-final *le* does not always function as a perfective marker.

Studies on the L1 acquisition of aspect marking in Mandarin suggest that monolingual speakers of Mandarin are able to acquire how the perfective and durative aspect markers *le/zhe* interact with verbs/predicates of different lexical aspects at an early age. With respect to the perfective aspect marker *le*, studies reported that Mandarin monolinguals started to produce *le* more and more frequently after the age of 1;7, and reached an adult-like performance in correctly associating *le* with verbs of different lexical aspect around 3;5 (Chen & Shirai 2010; Erbaugh 1992). Meanwhile, the production of the durative aspect marker *zhe* did not emerge in these monolinguals' speech until they reached the age of 3;0, but as long as these monolinguals started to produce *zhe*, they made very few errors in associating *zhe* with verbs of different lexical aspect (Jin & Hendriks 2003). Interestingly, Chen & Shirai (2010) observed that one 3-year-old child produced *zhe* with ACHs more frequently (6 out of 22 instances containing *zhe*) than her peer (1 out of 18 instances containing *zhe*). This pattern seemed to mirror the input from her caregivers, who continuously produced such a combination in speech.[5] This finding suggests that the quality of L1 input has an impact on monolingual children's acquisition of aspect marking in Mandarin; based on this fact, it is reasonable to speculate that the quality of HL input also has an impact on Mandarin HSs' acquisition of aspect marking, as these HSs acquire Mandarin as an L1. Therefore, if an HS of Mandarin showed non-monolingual-like knowledge of aspect marking, it is possibly due to that s/he has received input of changed quality – of course there will be other possible explanations for this phenomenon, but this falsifiable explanation is worth looking at, as it will help us to further understand whether and how input would affect the acquisition of aspect marking in a HL.

---

4. Not including "those indicating relations, psychological sensations, and adjectival verbs indicating personal properties (i.e. quality verbs)", as noted in Xiao & McEnery (2004: 189).

5. Unfortunately, the authors did not provide any example of ACH+*zhe* sentences produced by the parents, or the dialect background of the participants. Therefore, it is unclear whether this use of ACH+*zhe* is a mistake or an instance of language variation.

2.2   Studies on aspect marking in HL Mandarin

As indicated in the introduction, currently there seems to be no systematic study on L1 attrition of aspect marking in Mandarin. However, a few studies have looked into the acquisition/attrition of perfective aspect marking among young HSs of Mandarin.

Ming and Tao (2008) developed a corpus based on the compositions and essays written by undergraduate students enrolled in an elementary HL program in the US, and examined the HSs' use of the perfective *le* in this corpus. The researchers claimed that, out of the 1,217 instances containing *le*, the HSs "overused" *le* 47 times and "underused" it 137 times; therefore, these Mandarin HSs tended to "underuse" perfective aspect markers in obligatory contexts. However, the analysis was over-simplified and could not provide meaningful insights into Mandarin HSs' knowledge of aspect marking. On the one hand, the researchers classified the instances of *le* as grammatical, "overused" or "underused" based on their own intuition, making the classification unreliable. On the other hand, the researchers did not distinguish between verb-final *le* and sentence-final *le*, nor did they provide enough details about what kinds of errors the HSs had made (e.g. associating *le* with SEMs), so it remains unclear what problems those HSs had with producing *le* in written form. Moreover, as the corpus used in this study did not contain any natural speech from the HSs, one can hardly determine whether the HSs had developed non-monolingual-like knowledge about aspect marking, or whether they just had problems with developing monolingual-like literacy.

Using a story-retelling task, a written cloze test and an oral sentence completion task, Jia and Bayley (2008) studied the oral and written production of *le* by 36 Mandarin HSs from a HL school in the US, with ages ranging from 5 to 15. Out of these 36 HSs, half were born in the US or moved to the US before the age of 5, and the rest was born in China and moved to the US after the age of 6. According to the authors, the China-born HSs outperformed their US-born peers in all of the three tasks. Within the China-born group, the 5 to 10-year-old HSs outperformed those 11 to 15-year-old HSs in correctly supplying the obligatory *le* in the story-retelling and sentence completion tasks, while the older HSs outperformed the younger ones in the written cloze test; by contrast, with increasing age, the US-born HSs' performance in the story-retelling task did not change significantly, although their performance in the other two tasks changed in a pattern similar to the China-born HSs'. Unfortunately, although this study selected verbs of different lexical aspects, it did not examine whether the interaction between lexical and grammatical aspect had an effect on the participants' production of *le*, so it remains unknown whether the HSs had a global problem with using *le*, or specific problems with producing *le* with verbs of certain lexical aspects.

Shi (2011) studied the oral production of aspect markers by testing six adult Mandarin HSs in the Netherlands, whose age of onset of L2 immersion varied from 2;6 to 8;5. She used a video elicitation task, which required the participants to watch video clips and then describe what they saw in each clip. The researcher found that, comparing to the late bilinguals who were exposed to L2 Dutch after the age of 6, the early bilinguals who were exposed to L2 Dutch before or at the age of 4 were less likely to produce aspect markers. Moreover, she also observed that the early bilinguals tended to have difficulties with correctly producing imperfective aspect markers (especially the progressive *zai*) rather than perfective aspect markers.

The above studies have demonstrated that Mandarin HSs can have difficulties with aspect marking, yet none of them has investigated the possible causes of this phenomenon. As discussed in the introduction, it is possible that these HSs had received input of changed quality from the first generation speakers and thus developed non-monolingual-like knowledge about aspect marking. In order to verify this possibility, it is necessary to examine whether first generation overseas Mandarin speakers' knowledge of aspect marking is subject to attrition. In order to do so, we will test the grammatical knowledge of perfective and durative aspect marking in a group of Mandarin speakers with L2 English.

3. The present study

As demonstrated in the last section, the L1 acquisition of aspect marking in Mandarin can be influenced by the quality of L1 input during childhood, and young HSs seem to deviate from monolinguals in producing aspect markers in Mandarin. However, the previous studies on HL Mandarin only focused on young HSs of Mandarin and have not yet examined the HL input they received, so much is left unknown about the relationship between such deviations and the quality of input from the first generation.

Aiming to explore whether Mandarin HSs are likely to receive HL input of aspect marking with changed, non-monolingual-like quality, the present study will attempt to address an empirical research question: Is there any attrition of perfective and durative aspect marking among adult L1 Mandarin speakers who moved to the UK after puberty and have lived there for an extended period?

In order to answer this question, we will use an acceptability judgment task to test if the target participants would show attrition in perfective and durative aspect marking at the level of representation. As this study is a small scale pilot study, we have not been able to look into the production of perfective and durative aspect markers, which could directly unravel properties of the HL input containing aspect marking. However, it should be noted that if the participants would

show attrition in perfective and durative aspect marking in terms of grammatical knowledge, it is likely that they would deviate from monolinguals in producing these aspect markers as well; such production could further have an effect on the HL acquisition of aspect marking. Therefore, this study can still indirectly provide insights into whether Mandarin HSs are likely to receive HL input of changed, non-monolingual-like quality.

With respect to the prediction, we will follow the Interface Hypothesis (Sorace 2011), and hypothesize that our participants will not show attrition in perfective and durative aspect marking. The Interface Hypothesis suggests that, comparing to the language structures involving an "internal" interface between syntax and language-internal modules (e.g. syntax – lexicon), those structures involving an "external" interface between syntax and other cognitive domains (e.g. syntax – pragmatics) are more difficult to be acquired completely and more vulnerable to L1 attrition; therefore, near-native L2 learners and L1 attriters are more likely to show optionality and indeterminacy for structures at the "external" interface than those at the "internal" interface (see Domínguez 2013 for a critical review of this hypothesis). As will be shown in the Methodology section, the target sentences used in this study mainly involve interactions between syntax and lexicon, rather than interactions between syntax and discourse or pragmatics. Therefore, the language structure investigated here should be at the "internal" interface and thus resistant to L1 attrition.

4.   Methodology

4.1   Participants

Two groups of Mandarin speakers participated in this study. The experimental group consisted of 14 Mandarin-English bilingual speakers who had lived in the UK for at least seven years. All these bilinguals completed primary education before moving to the UK, and they were all literate in Mandarin. At the time of testing, these participants had completed secondary and/or higher education in the UK, and they were either university students or working professionals. For this group, the mean age was 33.07 (SD = 13.13), the mean age of arrival in the UK was 19.14 (SD = 7.13), and the mean length of residence in the UK was 13 years (SD = 7.46). Their self-reported proficiency scores for Mandarin and English were 4.86 (SD = 0.36) and 4.21 (SD = 1.05) respectively (1 = beginner, 5 = native command).

The control group consisted of 23 Mandarin monolinguals from Mainland China. At the time of testing, all these monolinguals had completed higher education, and none of them had lived in any country other than China. Moreover, their

contact with non-Chinese languages (e.g. English) after finishing higher education was minimal. The mean age of these speakers was 32.65 ($SD = 8.03$), and their self-reported proficiency scores for Mandarin and English were 5 ($SD = 0$) and 1.70 ($SD = 0.63$) respectively. In the rest of this paper, we will use "UK group" to refer to the Mandarin-English bilinguals in the UK, and "China group" to refer to the monolingual Mandarin speakers in China.

4.2   Material

This study used an acceptability judgment task to assess the participants' knowledge about the interaction between the six types of lexical aspect and the perfective and durative aspect markers. In this task, the participants read short sentences which either contained *le* or *zhe*, and judged the acceptability of the sentences using a Likert scale, with −2 being completely unacceptable, 0 being uncertain, and 2 being completely acceptable. The participants were tested in a one-to-one setting, and they finished this task using pencil and paper.

The stimuli were created by following a certain procedure. First, three verbs/ predicates of each lexical aspect[6] were selected from the examples provided in Xiao & McEnery (2004), so that a total of 18 verbs was chosen. After that, a short sentence was created for each verb. Then two versions of each sentence were created by adding *le* and *zhe* at the post-verbal position. By applying these steps, overall 36 stimuli sentences were created. Finally, the stimuli sentences were mixed with 27 fillers in pseudo-randomized order. In order to ensure that the participants' perception of these sentences was not affected by discourse/pragmatic factors, no context was provided with these sentences. Some examples of the stimuli sentences are given in Table 3.

As discussed in Section 2, Xiao and McEnery (2004) argued that *zhe* tended to co-occur with ACTs, ACCs, ILSs and SLSs, but not with SEMs or ACHs; meanwhile, *le* tended to co-occur with ACHs, ACCs, ACTs, but not with SEMs, ILSs or SLSs. Furthermore, they argued that, while *zhe* was strictly incompatible with ACHs or ILSs indicating personal properties, *zhe* could grammatically co-occur with SEMs and trigger an iterative reading. With respect to *le*, they argued that *le* could, although less frequently, grammatically co-occur with SEMs, ILSs and SLSs (for a detailed discussion see Xiao & McEnery 2004: 100–113, 188–194). Here we will follow their arguments and expect our participants to show different levels of acceptability for the stimuli sentences. More specifically, we expect our participants

---

6.   For the ILS predicates, we selected three ILSs indicating personal properties. These ILSs should be compatible with *le* but not *zhe* if we follow Xiao & McEnery (2004), see Section 2.1.

Table 3. Exemplar sentences from the acceptability judgment task

Type	Example
ILS+*le/*zhe*	*Ta pang le/zhe.* He fat PERF/DURA "He has become fat./He is being fat."
SLS+*le/zhe*	*Ta mang le/zhe.* He busy PERF/DURA "He has become busy./He is being busy."
ACT+*le/zhe*	*Ta chi le/zhe fan.* He eat PERF/DURA meal "He has had a meal./He is having a meal."
SEM+*le/zhe*	*Ta ke le/zhe sou.* He cough PERF/DURA cough "He coughed./He is coughing."
ACC+*le/zhe*	*Ta chi le/zhe yiwan fan.* He eat PERF/DURA one-bowl meal "He has eaten a bowl of rice./He is eating a bowl of rice."
ACH+*le/*zhe*	*Ta dao le/zhe zheli.* He arrive PERF/DURA here "He has arrived./He is arriving."

to accept the sentences containing ACC+*le/zhe*, ACT+*le/zhe*, SLS+*le/zhe*, SEM+*le/zhe*, ILS+*le* and ACH+*le*, but reject the ILS+*zhe* and ACH+*zhe* sentences. As we predicted no difference between the UK and the China group, it is expected that both groups will show similar levels of acceptance for each type of combination of lexical and grammatical aspect.

5. Results and analysis

The results for the UK and China groups are presented in Table 4. As the raw scores suggest, both the UK and the China group were inclined to reject the ACH+*zhe* and ILS+*zhe* sentences, while accepting the others; this pattern was consistent with what we expected. In order to examine if there was a statistically significant difference between the UK and the China group, we fitted a linear mixed-effects model for each type of verb/predicate+*le/zhe* combination. For this analysis, the raw scores were transformed into z-scores before being used as the dependent variable. In these models, Group was used as the predictor (China vs UK group; China group coded as 0.5, UK group as −0.5), and Subject and Item were treated as the random factors. Both Subject and Item had random intercepts, and Item had random slopes

Table 4. Mean acceptability scores (SDs) for the target sentences containing *le/zhe* + verbs of different lexical aspect

	ACC+*le*	ACC+*zhe*	ACH+*le*	ACH+*zhe*	ACT+*le*	ACT+*zhe*
UK Group	2 (0)	1.17 (0.95)	1.81 (0.31)	−1.88 (0.21)	1.62 (0.89)	1.48 (0.89)
China Group	1.93 (0.28)	0.88 (1.02)	1.84 (0.39)	−1.74 (0.45)	1.87 (0.28)	1.43 (0.87)

	ILS+*le*	ILS+*zhe*	SEM+*le*	SEM+*zhe*	SLS+*le*	SLS+*zhe*
UK Group	0.79 (0.81)	−1.62 (0.54)	0.71 (0.94)	0.79 (0.82)	0.86 (0.5)	1.07 (1.02)
China Group	1.09 (0.82)	−1 (1.03)	1.03 (0.63)	0.97 (0.92)	1.04 (0.51)	1.51 (0.66)

for the fixed effect of Group. This analysis only revealed a marginally significant Group effect on the acceptability scores for the ILS+*zhe* sentences (Estimate = 0.39, $t = 1.89$, $SE = 0.21$, $p = 0.08$), but no significant Group effect for sentences of any other type.

Although the statistical analysis revealed a marginally significant effect of Group on the acceptability scores for the ILS+*zhe* sentences, which suggests that the bilinguals were more likely to reject the ILS+*zhe* sentences than the monolinguals, this effect does not indicate that the bilinguals had non-monolingual-like grammatical knowledge about the interaction between ILSs and *zhe*, because both groups tended to reject the ungrammatical ILS+*zhe* sentences. The high level of variation within the China group (as indicated by the SD) suggests that the China group might have treated one of the ILS+*zhe* sentences differently, or that some participants in the China group behaved differently from the others in rating these sentences. There are at least two possible explanations for such a behavior. First, it may be the case that some monolinguals mistook *zhe* for an intensifier, such as *zhe ne*, and accepted the ungrammatical ILS+*zhe* sentences (see Li & Thompson 1989: 222 for a discussion about *zhe ne*). Second, it may be caused by the fact that some monolinguals had different dialect backgrounds and tended to treat ILS+*zhe* differently. Therefore, we examined each individual participant's data and found two participants consistently accepting all three ILS+*zhe* sentences, as well as three participants accepting two of the three ILS+*zhe* sentences. Unfortunately, since we do not have their dialect background, it is impossible to do any further analysis. It should be noted, though, that we are mainly interested in whether the UK group would deviate from the monolinguals in incorrectly accepting ungrammatical sentences or rejecting grammatical ones, rather than how monolinguals would perform in this task, so this issue will not be further discussed here.

6.   General discussion

The results of this study suggest that the Mandarin speakers in the UK still have monolingual-like knowledge about the interaction between lexical aspect and the grammatical aspect markers *le/zhe*, since they behaved similarly in rating the target sentences of most types. Although we observed one marginally significant between-group difference in rating the ungrammatical ILS+*zhe*, this difference could not support the idea that the UK group showed attrition of the grammatical knowledge about the interaction between ILSs and *zhe*: Both the UK and the China group tended to reject these sentences, and neither group tended to incorrectly accept them. Therefore, it is reasonable to conclude that the UK group did not show attrition in the interaction between lexical aspect and *le/zhe* in terms of grammatical knowledge. This finding is also consistent with the Interface Hypothesis, which predicts that no attrition occurs in language structures at the "internal" interfaces.

Given that we had a small sample and only tested three predicates of each lexical aspect in this study, we cannot reach a substantial conclusion solely based on the findings of this study. However, future research should look more into this issue. Moreover, while our findings did not confirm attrition among the Mandarin speakers in the UK, they did not exclude the possibility that these speakers would deviate from the monolinguals in producing aspect markers either. For example, it is possible that overseas Mandarin speakers would avoid producing *zhe* with verbs/predicates of certain lexical aspects (e.g. SEMs, as SEM+*zhe* is infelicitous) in their speech; if this is the case, Mandarin HSs are still likely to receive HL input of changed quality. Therefore, it is necessary for future studies to look into the production of aspect marking, and directly compare Mandarin HSs with their parents.

In addition, it would also be interesting to look into a few other issues in the L1 attrition and HL acquisition of aspect marking in Mandarin. Regarding L1 attrition, it would be worth investigating whether overseas Mandarin speakers would show different levels of attrition for perfective and imperfective aspect marking respectively, and whether L1 attrition in aspect marking could happen when specific contexts were provided: In this case, aspect marking involves interactions between syntax, lexicon and discourse, and is therefore at the "external" interface. With respect to HL acquisition, future research could look into the other possible causes which lead to deviant outcomes of acquiring aspect marking, such as cross-linguistic influence; future research could also test how HSs perceive and interpret the temporal information encoded in bare verbs/predicates, and study whether the HSs' problems with using aspect markers were caused by non-monolingual-like representation of lexical aspect encoded in verbs/predicates, rather than the representation of grammatical aspect markers.

7.   Conclusion

This study investigated whether adult L1 Mandarin speakers who have lived in the UK for an average of 13 years would show attrition of perfective and durative aspect marking. An acceptability judgment task was employed to examine their knowledge of the target grammatical structures. A comparison between the Mandarin speakers in the UK and those in China revealed that these two groups did not differ in their knowledge about perfective or durative aspect marking.

Although this study has limitations regarding its findings and sample size, it has served its purpose of exploring the under-researched L1 attrition of aspect marking among adult overseas Mandarin speakers. Hopefully, future research on the same topic will help us to reach a better understanding about L1 attrition of aspect marking in Mandarin, as well as how L1 attrition would affect the quality of HL input and the acquisition of HLs.

References

Aalberse, S., Zou, Y. & Andringa, S. 2017. Extended use of demonstrative pronouns in two generations of Mandarin Chinese speakers in the Netherlands: Evidence of convergence? In *Crosslinguistic Influence in Bilingualism: In Honor of Aafke Hulk*, [Studies in Bilingualism 52], E. Blom, L. Cornips & J. Schaeffer (eds), 25–48. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.52.03aal

Brehmer, B. & Kurbangulova, T. 2016. Effects of parental input on heritage language development: A comparison across linguistic categories. Paper presented at the Workshop on Heritage Language Acquisition, The Arctic University Tromsø, 19./20.09.2016.

Chen, J. & Shirai, Y. 2010. The development of aspectual marking in child Mandarin Chinese. *Applied Psycholinguistics* 31(1): 1–28. https://doi.org/10.1017/S0142716409990257

Comrie, B. 1976. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge: CUP.

Ding, T. 2013. New type of learner emerging: Understanding learners of Chinese as a Heritage Language. *Cambridge Journal of China Studies* 8(2): 49–61.

Domínguez, L. 2013. *Understanding Interfaces: Second Language Acquisition and Native Language Attrition of Spanish Subject Realization and Word Order Variation* [Language Acquisition and Language Disorders 55]. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.55

Duff, P. & Li, D. 2002. The acquisition and use of perfective aspect in Mandarin. In *The L2 Acquisition of Tense-Aspect Morphology* [Language Acquisition and Language Disorders, 27], R. Salaberry & Y. Shirai (eds), 417–454. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.27.17duf

Erbaugh, M. S. 1992. The acquisition of Mandarin. *The Crosslinguistic Study of Language Acquisition*, Vol. 3, D. I. Slobin (ed.), 373–455. Hillsdale NJ: Lawrence Erlbaum Associates.

Hui, S. 2012. First Language Attrition of Chinese Reflexives. PhD dissertation, The Chinese University of Hong Kong.

Jia, L. & Bayley, R. 2008. The (re) acquisition of perfective aspect marking by Chinese heritage language learners. In *Chinese as a Heritage Language: Fostering Rooted World Citizenry*, A. W. He & Y. Xiao (eds), 205–222. Honolulu HI: University of Hawai'i Press.

Jin, L. & Hendriks, H. 2003. The development of aspect marking in L1 and L2 Chinese. *Working Papers on Applied Linguistics* 9: 69–100.

Klein, W., Li, P., Hendriks, H. & Language, S. N. 2000. Aspect and assertion in Mandarin Chinese. *Natural Language & Linguistic Theory* 18(4): 723–770.
https://doi.org/10.1023/A:1006411825993

Kupisch, T. 2013. A new term for a better distinction? A view from the higher end of the proficiency scale. *Theoretical Linguistics* 39(3–4): 203–214.

Li, C. & Thompson, S. 1989. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley CA: University of California Press.

Li, G. 2006. Biliteracy and trilingual practices in the home context: Case studies of Chinese-Canadian children. *Journal of Early Childhood Literacy* 6(3): 355–381.
https://doi.org/10.1177/1468798406069797

Liu, M. 2015. Tense and aspect in Mandarin Chinese. *The Oxford Handbook of Chinese Linguistics*, W. S.-Y. Wang & C. Sun (eds), 274–289. Oxford: OUP.

Luo, H. 2015. Chinese language learning anxiety: A study of heritage learners. *Heritage Language Journal* 12(1): 22–47.

Ming, T. & Tao, H. 2008. Developing a Chinese heritage language corpus: Issues and a preliminary report. In *Chinese as a Heritage Language: Fostering Rooted World Citizenry*, A. W. He & Y. Xiao (eds), 167–187. Honolulu HI: University of Hawai'i Press.

Montrul, S. 2008. *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor* [Studies in Bilingualism 39]. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.39

Montrul, S. 2016. *The Acquisition of Heritage Languages*. Cambridge: CUP.
https://doi.org/10.1017/CBO9781139030502

Pascual y Cabo, D. & Rothman, J. 2012. The (il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied Linguistics* 33(4): 450–455.
https://doi.org/10.1093/applin/ams037

Peck, J., Lin, J. & Sun, C. 2013. Aspectual classification of Mandarin Chinese verbs: A perspective of scale structure. *Language and Linguistics* 14(4): 663–700.

Polinsky, M. 2011. Reanalysis in adult heritage language. *Studies in Second Language Acquisition* 33(2): 305–328. https://doi.org/10.1017/S027226311000077X

Rothman, J. 2007. Heritage speaker competence differences, language change, and input type: Inflected infinitives in Heritage Brazilian Portuguese. *International Journal of Bilingualism* 11(4): 359–389. https://doi.org/10.1177/13670069070110040201

Rothman, J. & Treffers-Daller, J. 2014. A prolegomenon to the construct of the native speaker: Heritage speaker bilinguals are natives too! *Applied Linguistics* 35(1): 93–98.
https://doi.org/10.1093/applin/amt049

Schmid, M. S. 2016. First language attrition. *Language Teaching* 49(2): 186–212.
https://doi.org/10.1017/S0261444815000476

Schmid, M. S. & Köpke, B. 2017. The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism* 7(6): 637–667.
https://doi.org/10.1075/lab.17058.sch

Shi, M. 2011. Incomplete Knowledge of Aspect in Heritage Speakers of Mandarin Chinese in the Netherlands. MA thesis, Radboud University, Nijmegen.

Smith, C. S. 1997. *The Parameter of Aspect*. Dordrecht: Kluwer.
https://doi.org/10.1007/978-94-011-5606-6

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

Valdés, G. 2005. Bilingualism, heritage language learners, and SLA research: Opportunities lost or seized? *The Modern Language Journal* 89(3): 410–426.
https://doi.org/10.1111/j.1540-4781.2005.00314.x

Vendler, Z. 1967. *Linguistics in Philosophy*. Ithaca NY: Cornell University Press.

Wiedenhof, J. 2015. *A Grammar of Mandarin*. Amsterdam: John Benjamins.
https://doi.org/10.1075/z.197

Xiao, R. & McEnery, T. 2004. *Aspect in Mandarin Chinese: A Corpus-based Study* [Studies in Language Companion Series 73]. Amsterdam: John Benjamins.
https://doi.org/10.1075/slcs.73

Author index

Language index

Subject index

Heritage speakers are a fascinating group of bilinguals with a unique profile. Living abroad as immigrants of the second generation, they speak the language of their own speech community (the heritage language) at home, and the societally dominant language in most other domains. What exactly they know about their heritage language continues to fascinate the research community as well as teachers and other practitioners working with this group. The different contributions cover a large variety of studies into heritage languages spoken in Europe and North America (including Chinese, Norwegian, Russian, Spanish and Turkish). The volume makes a key contribution to the description and explanation of variability in the outcomes of heritage language acquisition, taking into account a wide range of factors which impact on language acquisition. As comparisons are frequently made with monolinguals and foreign language learners, the volume is also highly relevant for researchers working in monolingual language acquisition and foreign language learning and teaching.

JOHN BENJAMINS PUBLISHING COMPANY