Article

Morphosyntactic Development in First Generation Arabic—English Children: The Effect of Cognitive, Age, and Input Factors over Time and across Languages

Johanne Paradis, Adriana Soto-Corominas, Evangelia Daskalaki, Xi Chen and Alexandra Gottardo

Article

Morphosyntactic Development in First Generation Arabic—English Children: The Effect of Cognitive, Age, and Input Factors over Time and across Languages

Johanne Paradis [1,*], Adriana Soto-Corominas [2], Evangelia Daskalaki [1], Xi Chen [3] and Alexandra Gottardo [4]

1. Department of Linguistics, University of Alberta, Edmonton, AB T6G 2E7, Canada; daskalak@ualberta.ca
2. Department of Applied Linguistics, Universitat Internacional de Catalunya, 08017 Barcelona, Spain; asotoc@uic.cat
3. Department of Applied Psychology and Human Development, Ontario Institute for Studies in Education, University of Toronto, Toronto, ON M5S 1V6, Canada; xi.chen.bumgardner@utoronto.ca
4. Department of Psychology, Wilfrid Laurier University, Waterloo, ON N2L 3C5, Canada; agottardo@wlu.ca
* Correspondence: jparadis@ualberta.ca

Abstract: This longitudinal study examined morphosyntactic development in the heritage Arabic-L1 and English-L2 of first-generation Syrian refugee children (mean age = 9.5; range = 6–13) within their first three years in Canada. Morphosyntactic abilities were measured using sentence repetition tasks (SRTs) in English and Syrian Arabic that included diverse morphosyntactic structures. Direct measures of verbal and non-verbal cognitive skills were obtained, and a parent questionnaire provided the age at L2 acquisition onset (AOA) and input variables. We found the following: Dominance in the L1 was evident at both time periods, regardless of AOA, and growth in bilingual abilities was found over time. Cognitive skills accounted for substantial variance in SRT scores in both languages and at both times. An older AOA was associated with superior SRT scores at Time−1 for both languages, but at Time-2, older AOA only contributed to superior SRT scores in Arabic. Using the L2 with siblings gave a boost to English at Time−1 but had a negative effect on Arabic at Time-2. We conclude that first-generation children show strong heritage-L1 maintenance early on, and individual differences in cognitive skills have stable effects on morphosyntax in both languages over time, but age and input factors have differential effects on each language and over time.

Keywords: child bilingualism; second language acquisition; heritage language acquisition; morphosyntax; individual differences

1. Introduction

Much recent research has focused on the unique morphosyntactic properties of languages when they are acquired as a first language (L1) in a heritage context by bilingual children, who typically become dominant in the societal language, their second language (L2). Nevertheless, few studies have tracked the morphosyntactic development in both languages of heritage-L1, English-L2 bilingual children using a longitudinal design. Those studies that have done so have mainly focused on simultaneous bilinguals during the preschool years (e.g., Hoff et al. 2012; Hoff and Ribot 2017; Silva-Corvalán 2014; Yip and Matthews 2007). Therefore, there is limited research on the trajectories of morphosyntactic development in both languages in school-age sequential bilinguals, especially documenting the shift in dominant language from the heritage-L1 to the L2. Retrospective studies with adults show that a shift in dominance to the L2 is the norm for child bilinguals from minority backgrounds in North America (e.g., Carreira and Kagan 2011; Montrul 2016), but the age when this commonly occurs and the impact on heritage-L1 maintenance is not yet well documented through prospective, longitudinal research. In addition, studies of heritage-L1 morphosyntactic acquisition to date have mainly been conducted with second-generation children or adults, i.e., individuals who were born in the host country, or with

mixed-generation samples (see Montrul 2016, for review). There is comparatively less research documenting the development of the heritage-L1 alongside the L2 in first-generation child migrants at the point where dual language learning begins. Thus, examining the prospective development of the heritage-L1 and L2 over time in first-generation children would contribute further to our knowledge of dominant language shift.

There has been much recent research on the role of child and environmental factors in determining the rate of L2 morphosyntactic acquisition, and to a lesser extent, heritage-L1 morphosyntactic acquisition (see Armon-Lotem and Meir 2019, for review). Existing research on sources of individual differences in morphosyntactic acquisition suggests that cognitive abilities, age of L2 acquisition onset (AOA) and language input factors can impact the heritage-L1 and the L2, with AOA and some input factors showing differential effects for the heritage-L1 and L2. However, few studies have examined the impact of these factors on the two languages of the same children using a longitudinal design. Doing so would provide greater insight into how morphosyntactic acquisition in the two languages is shaped by these factors, as well as the contribution of these factors to dominant language shift.

In 2015, the Canadian government launched an unprecedented program to resettle Syrian refugee families and to date, over 70,000 refugees have arrived in Canada (Immigration, Refugees and Citizenship Canada (IRCC 2020)). This migration has provided the opportunity to examine the bilingual development of first-generation children who arrived as a cohort and speak the same heritage-L1. Child refugees like those in the present study often experience adversity that could directly impact their overall development, including language development, for example, interrupted schooling in their L1, displacement and frequent transitions, exposure to trauma and poverty (Graham et al. 2016; Sirin and Rogers-Sirin 2015; Kaplan et al. 2016). Therefore, the heritage-L1 of these first-generation children could potentially be more at risk than would be the case for child migrants who arrive to a host country with age-appropriate education levels in their L1, no significant adversity, and greater socioeconomic advantages.

The present study investigated the morphosyntactic development of the English-L2 and Arabic heritage-L1 by Syrian refugee children newly arrived in Canada at two time points within their first three years of residency. Our objectives were twofold: First, we sought to determine the developmental trends for L2 and L1 morphosyntax over this early time period in dual language development, more particularly to look for dominant language shift. Second, we sought to determine how cognitive, age, and input factors influenced children's morphosyntactic abilities, in particular whether they had differential effects on the L2 versus the L1 and at Time 1 versus Time 2.

1.1. Dominant Language Shift in Bilingual Acquisition

Heritage-L1 children, regardless of immigration depth/generation, are expected to shift to dominance in the societal-L2, when educated in the L2, at some age before adulthood in the North American context (Carreira and Kagan 2011; Montrul 2016). Nevertheless, there has been limited research examining dominant language shift prospectively and existing research shows some variable and conflicting findings.

Regarding simultaneous bilinguals, Silva-Corvalán's (2014) study of two Spanish–English bilingual siblings living in the United States revealed that the children showed balanced bilingual development up until the age of 3. Based on a larger sample, Hoff and colleagues found that Spanish–English bilingual children can show dominance in English vocabulary and morphosyntactic development before 2 1/2 years of age (Hoff et al. 2012) and slower growth in Spanish than English for expressive vocabulary from 2 1/2 to 5 years of age (Hoff and Ribot 2017). Hammer et al. (2011) found that shift to English dominance was dependent on language use at home among Spanish–English bilinguals. Specifically, they found that children who spoke mainly English at home shifted to English dominance between the ages of 3 1/2 and 5 years while attending an English-only Head Start program,

but children in the program who spoke mainly Spanish at home had not yet shifted to English dominance by age 5.

Unlike simultaneous bilinguals, there is a wider range with respect to the age when sequential bilingual children begin to learn the majority-L2, which in turn can affect dominant language shift. Kohnert and colleagues examined the lexical processing skills in both Spanish and English by bilingual children and adolescents, from 5 to 16 years of age, in Southern California (Kohnert and Bates 2002; Kohnert et al. 1999). The participants' lexical processing skills were stronger in Spanish initially and only shifted to English in late childhood/early adolescence. In contrast, studies examining dominant language shift for heritage languages that have smaller speaker communities than Spanish in California have found that middle childhood (around 7–9 years of age) is a common age for a child's dominant language to shift to the majority-L2. For example, a cross-sectional study examining vocabulary skills in the heritage-L1 (Mandarin, Cantonese, or Korean) of English-L2 learners in the United States found stronger skills in English than the heritage language starting from age 7 (Jia et al. 2014). Pham and Kohnert (2014) followed a group of Vietnamese-L1 English-L2 children in bilingual education programming from 7 to 11 years of age. They found that the dominant language shifted from Vietnamese to English around age 8 for expressive vocabulary skills but did not shift for receptive skills by 11 years of age. However, the rate of growth was steeper for English than for Vietnamese during this time. Unlike participants in G. Jia et al. (2014), these children were exposed to rich input in the L1 through schooling, which seemed to delay dominance shift in receptive vocabulary skills.

The research we have reviewed thus far was based mainly on second-generation or mixed samples of participants, and there are fewer studies with first-generation child bilinguals. First-generation children could be expected to show superior heritage-L1 maintenance because they have had a longer period of monolingual development, monolingual language environments, and possibly some schooling in the L1. One study of first-generation participants, Jia and Aaronson (2003), examined the Mandarin heritage-L1 and English-L2 development in new arrivals over a 3-year period of residency in the United States. Language preference, use, and proficiency in both languages were examined, and they found that participants who were younger at arrival (<9 years) shifted their dominance to the L2 by the end of this 3-year period, while the participants who arrived as adolescents (>12 years) did not. Limitations of this study include a small sample size (*n* = 10), and a mismatch between measures of proficiency between the L2 and L1 (grammaticality judgement task in the L2; parent proficiency ratings in the L1). Another relevant study, Hamann et al. (2020), consisted of a comparison between first-generation and second-generation Arabic heritage-L1 German-L2 children aged 6–12 years old. The first-generation children were recent arrivals from refugee backgrounds and had 18 months of German-L2 training through special classes for newcomers. The second-generation group were more likely to be L2 dominant than the new arrivals since they had stronger L2 skills on a sentence repetition task (SRT) in German than in Arabic. Their results suggest that a shift to L2 dominance is later for first-generation migrant children.

In sum, the existing research on dominant language shift in bilingual children shows that the age at which this occurs can be variable and modulated by AOA and input factors. No existing study has examined morphosyntax using the same measure in both languages over time in school age bilinguals to examine dominant language shift. Accordingly, this was one goal of the present study.

1.2. The Influence of Cognitive Abilities on L2 and Heritage-L1 Development

Cognitive capacities that are implicated in language learning, such as short-term memory, working memory, and analytic reasoning skills, have been found to predict individual variation in children's bilingual development. This relationship is likely due to these cognitive abilities being key components of language aptitude, a robust predictor of success in L2 learning, among adults in particular (Paradis 2011). In Paradis (2011) and Paradis et al. (2017), both verbal short-term memory and non-verbal analytic reasoning

were found to be strong predictors of lexical, morphological, and syntactic abilities in school-age English-L2 children in Canada. R. Jia and Paradis (2020) also found that individual differences in verbal memory were associated with bilingual children's syntactic abilities in their Mandarin heritage-L1. Pham and Tipton (2018) found that superior verbal short-term memory skills were associated with larger vocabularies in both the Vietnamese heritage-L1 and English-L2 of school age bilingual children in the United States (see also, Collins et al. 2014 for Spanish–English bilinguals). Regarding studies with Arabic heritage-L1 bilinguals, Paradis et al. (2020a) reported a robust relationship between non-verbal analytic reasoning abilities and Arabic-L1 and English-L2 vocabulary knowledge and morphological abilities in school age bilinguals. Similarly, Hamann et al. (2020) found that differences in working memory abilities were predictive of differences in Arabic-L1 German-L2 children's performance on a sentence repetition task in both languages (but see Andreou et al. 2020 who did not find evidence of a relationship).

Since this growing body of research demonstrates the important role of cognitive capacities in bilingual development, we have included measures of cognitive capacities in the present study for our analyses of individual differences in morphosyntactic development.

1.3. The Influence of AOA on L2 and Heritage-L1 Development

On one hand, an earlier start to L2 learning is typically associated with more native-like competence in the L2, especially when comparing those who began to learn their L2 in childhood with those who learned the L2 as adults (Birdsong and Molis 2001; DeKeyser 2012; Johnson and Newport 1989). On the other hand, studies have shown that older AOA within the early childhood years is associated with more advanced L2 development in vocabulary and morphosyntax for children in L2-majority contexts, when the amount of L2 exposure has been controlled for (Chondrogianni and Marinis 2011; Golberg et al. 2008; Paradis 2011). However, Jia and Aaronson (2003) and Jia and Fuse (2007), who included participants with AOAs ranging from 5–16 years, found that the older-learner advantage was not borne out over time and younger arrivals surpassed the older arrivals in L2 abilities after 3 to 5 years of L2 exposure.

Regarding maintenance of the heritage-L1, younger AOA children have had a shorter period of being monolingual, which can be associated with higher rates of attrition and divergent acquisition patterns in child and adult heritage speakers (Albirini 2018; Jia 2008; Jia and Paradis 2015, 2020; see Montrul 2016 for review). For example, Mandarin–English bilingual children in Canada with an older AOA showed stronger Mandarin heritage-L1 narrative and complex syntax abilities than those with a younger AOA (Jia and Paradis 2015, 2020). Similarly, Meir, Walters, and Armon-Lotem (Meir et al. 2017) found that Russian–Hebrew bilingual children with an older AOA displayed higher morphosyntactic accuracy in their Russian heritage-L1 than in their Hebrew L2. Regarding Arabic heritage-L1 bilinguals, both Albirini (2018) and Paradis et al. (2020a) found that children with older AOAs had larger vocabularies and better morphosyntactic skills in Arabic than their peers who began learning the L2 at younger ages.

In sum, findings for AOA with respect to L2 acquisition suggest that there can be short-term benefits of older AOA but long-term benefits of younger AOA. In contrast, there is a consensus in findings for the heritage-L1: older AOA is more protective against attrition, divergent acquisition patterns, and by extension, early dominant language shift. Differential effects of AOA have not often been examined for both languages in the same children, and thus, we did so in the present study. We looked at the role of AOA with respect to shift in dominance to the L2 as well as a source of individual differences in morphosyntactic development of each language separately.

1.4. The Influence of Language Environment on L2 and Heritage-L1 Development

A robust finding in the literature is that greater quantity of language input is related to greater proficiency in bilingual children. The research also indicates that the sources and qualitative aspects of input, together with quantity, can have differential impacts on

the heritage-L1 and the majority-L2. The latter was a particular motivation for examining the relations between these factors and children's morphosyntactic development in both languages over time in the present study. We use "language environment" as a cover term for these various and interacting input factors.

1.4.1. Language Use at Home

Input in the L2 from parents who are less fluent/not native speakers is less supportive of English L2 development than input from fluent/native English-speaking parents in preschool- and school-age children (Chondrogianni and Marinis 2011; Hammer et al. 2012; Paradis 2011; Place and Hoff 2016; Sorenson Duncan and Paradis 2020a). In addition to fluency, the interlocutor also makes a difference. Rojas et al. (2016) found that stronger English-L2 expressive skills (mean utterance length and lexical diversity) in Spanish–English bilinguals were predicted by more interactions with siblings and peers (who used more English than parents). Similarly, Sorenson Duncan and Paradis (2020b) found that L2 input-output with older siblings positively influenced lexical and morphosyntactic skills in 5-year-old English L2 learners from diverse L1 backgrounds, but there was limited evidence that L2 input-output from mothers had an effect. Paradis et al. (2020a) also found that more sibling interaction in English was associated with larger English-L2 vocabularies and morphological abilities among Arabic heritage-L1 children (see also Hamann et al. 2020).

Use of the heritage-L1 at home could be even more predictive of its development since bilingual children have access to the majority L2 in the community and at school, but more restricted access to the heritage-L1 outside the home. Many studies have found a positive relationship between more heritage-L1 use at home among family members and proficiency in the heritage-L1 among preschool children (Altman et al. 2014; Hammer et al. 2012; Place and Hoff 2016; Prevoo et al. 2014), school-age children (Daskalaki et al. 2019; Daskalaki et al. 2020; Pham and Tipton 2018; Rojas et al. 2016; Sorenson Duncan and Paradis 2020a), and adolescents (Flores et al. 2017). Studies often find language choices among family members affect children's heritage-L1 and majority-L2 development in different ways. For example, Hammer et al. (2011) found that increases in the use of English at home in Spanish-speaking families when children entered an English Head Start program had no impact on children's English growth rate; whereas, it had a negative effect on children's Spanish growth rate (see also Rojas et al. 2016).

1.4.2. Richness of the Language Environment

Richness refers to the amount of diverse and complex input and output children experience, for example, the frequency of children's engagement in L1 or L2 print and audiovisual media, extra-curricular activities in the L1 or L2, and socializing with friends in the L1 or L2. Studies including a composite L2 richness variable have shown this to positively promote stronger L2 vocabulary, morphology, syntax and narrative skills (Govindarajan and Paradis 2019; Jia and Aaronson 2003; Jia and Fuse 2007; Paradis 2011; Paradis et al. 2017; Paradis et al. 2020a). Regarding home literacy practices in particular, the frequency of reading and other literacy-related activities in the L2 can enhance L2 development in school-age bilinguals (Kaltsa et al. 2019; Prevoo et al. 2014).

Like language use with family members, richness of the linguistic environment is possibly more vital to the acquisition of the heritage-L1 than to the L2 (Jia 2008). Jia and Aaronson (2003) found that greater richness of the L1 environment was positively correlated with parents' report on child and youth L1 proficiency. Greater frequency in an average week of media, extra-curricular activities, and socializing in Mandarin was associated with stronger Mandarin heritage-L1 narrative abilities in R. Jia and Paradis (2015). Pham and Tipton (2018) found a similar positive association between Vietnamese heritage-L1 richness at home and Vietnamese–English children's vocabulary in this language. Studies with youth and adults have found that being literate in the heritage-L1 and having access to written media in the heritage-L1 promotes long-term maintenance (Albirini 2014; Andreou et al. 2020; Bayram et al. 2017; Jia 2008).

1.4.3. Maternal Education

Maternal education is a broad language environment factor because it indexes greater quantity and quality of linguistic input to children as well as overall family environment and cultural capital (Prevoo et al. 2014). Higher maternal education is associated with greater proficiency in English among Spanish–English preschoolers (Place and Hoff 2016; Hammer et al. 2012; Rojas et al. 2016). Golberg et al. (2008) found that higher maternal education predicted larger L2 vocabularies across 5 time points in a 2-year longitudinal study on school-age, English-L2 children with diverse L1 backgrounds. Sorenson Duncan and Paradis (2020a) and Prevoo et al. (2014) found that the relationships between maternal education and children's L2 development were mediated by mothers' L2 fluency and mothers' engagement with home literacy activities, respectively. Turning to studies of Arabic heritage-L1 children, both Paradis et al. (2020a) and Hamann et al. (2020) found higher maternal education to be associated with stronger L2 lexical and morphosyntactic abilities.

In contrast, the connection between maternal education levels and bilingual children's heritage-L1 development has produced conflicting findings. There was no evidence for an impact of maternal education on the L1 (but there was on the L2) in Place and Hoff (2016), Prevoo et al. (2014), Rojas et al. (2016), and Hamann et al. (2020). In contrast, R. Jia and Paradis (2015) found that higher maternal education predicted stronger Mandarin heritage-L1 skills and Paradis et al. (2020a) found maternal education to be associated with stronger Arabic heritage-L1 skills. Armon-Lotem et al. (2011) found that maternal education positively influenced Russian heritage-L1 maintenance in Russian–German bilinguals in Germany, but not in Russian–Hebrew bilinguals in Israel. Sorenson Duncan and Paradis (2020a) reported an indirect effect of maternal education on L1 proficiency where the effect was mediated by the association between language of education and language choice (see also Prevoo et al. 2014).

1.5. The Present Study

The present study examined the English-L2 and Syrian Arabic heritage-L1 morphosyntactic acquisition in school age bilingual children who were first-generation migrants. We measured their morphosyntactic abilities using sentence repetition tasks in each language which varied in the complexity of the sentences, for example, stimuli were comprised of simple declaratives as well as passives and relative clauses. We assessed participants' morphosyntactic skills twice: once after a mean of two years of residency in Canada (Time-1; henceforth, T1) and again one year later (T2). We also included measures of cognitive abilities, verbal short-term memory, and nonverbal analytic reasoning, as well as an extensive parent questionnaire to gather information on AOA and language environment factors. We asked the following research questions:

RQ1: What developmental trends emerge in children's L1 and L2 between T1 and T2? Is there evidence of dominant language shift?

On one hand, first-generation child migrants could be expected to show strong maintenance of the heritage-L1 during the first 3 years of residency. On the other hand, many of these Syrian children had interrupted or no prior schooling in Arabic because of the war (Paradis et al. 2020a), and this could mean that L1 skills, especially for complex morphosyntax, could be weaker than expected. In addition, there is no consensus in the existing research on when children shift to the L2 as their dominant language, and research with first-generation children is sparse. We approached this question in two ways: First, we compared scores on the SRT in each language at T1 and T2 to determine whether, as a group, signs of superior English-L2 to Arabic-L1 skills were apparent. Second, because the children in this study varied in age, and thus AOA, we also examined the role of AOA (younger and older) in relative performance in each language across the two time periods.

RQ2: What is the association between cognitive abilities, AOA, and input factors to English-L2 and Arabic-L1 morphosyntax at T1 and at T2? Do associations differ by language (L1 vs. L2) or time (T1 versus T2)?

Based on prior research, we expected superior cognitive abilities to positively influence task performance in both languages, and at both time points. For AOA, on the one hand, younger AOA might be associated with superior L2 performance while older AOA would be associated with superior L1 performance. On the other hand, because the children are at the early stages of acquiring their L2, older AOA might be an advantage in performance on the task in both languages. Regarding differences over time, younger AOA could be associated with weaker L1 skills at T2. For input factors, expectations depended on the language, the type of factor and the time interval. We expected increased richness in the English and Arabic environment outside school to be associated with stronger outcomes in each language in general, but that this factor would be increasingly more important for the L1 overtime. By contrast, we expected differential effects for relative use of English and Arabic at home, with greater use of the L2 associated with better L2 performance but worse L1 performance, especially over time. Since maternal education level has been more reliably linked to the societal-L2 than to the heritage-L1 in prior research, we expected to also see differential effects in our data and at both time points.

2. Materials and Methods

2.1. Participants

This study includes 102 participants (49 females), all of them Syrian children with a refugee background. These 102 participants are part of the initial sample of 133 participants for an ongoing longitudinal study across three cities in Canada (cf. Paradis et al. 2020a). Only those participants who completed both the English and Syrian Arabic tests at both T1 and T2 are included here. Contributions to the attrition rates at T2 included: families moving away, families declining to participate at T2 (only $N = 2$), and issues with recording equipment for one of our measures at either T1 or T2.

At T1, these participants had a mean age of 9.5 years ($SD = 1.89$; range 6–13) and had lived in Canada for a mean of 24 months ($SD = 7.20$). T2 data was collected approximately 11 months later ($M = 10.65$; $SD = 2.39$). All participants and their parents were native speakers of Syrian Arabic. None of the participants had been exposed to English before their arrival to Canada. The children migrated with their families in 2015–2017, and began attending school in English 1.55 months ($SD = 1.70$) after arrival in Canada. Age of L2 acquisition was determined to begin at English schooling, since children entered schools relatively soon after arrival. Children were residing in one of three cities in Canada: Edmonton, Toronto, and Waterloo. Children were attending English medium schools, in mainstream classrooms with additional English second language instruction. How this instruction was delivered varied among schools and cities, but most children received English second language instruction through a pull-out method, i.e., children were given instruction one-on-one or in small groups outside of the regular class. More information about participant characteristics and family demographics is presented in the Results section.

2.2. Procedures

The present study was conducted in accordance with the Declaration of Helsinki and the protocol was approved by the university research offices in the pertinent universities in Edmonton, Toronto, and Waterloo. Parents gave informed consent to participate and children 8 years old or older also gave their assent, in accordance with ethical guidelines in place in Canada.

2.2.1. Alberta Language Environment Questionnaire-4 (ALEQ-4)

To obtain information on the child-level variables that might influence participants' performance in their two languages, we used the Alberta Language Environment Questionnaire-

4 (ALEQ-4; Paradis et al. 2020a). This questionnaire was delivered to parents in an interview format in Arabic by a native speaker of Arabic and gathered demographic information on both the family and the child (age, length of residency in Canada, English AOA, length of schooling in English and Arabic, and parental education), together with information on the amount of language use at home as well as the overall richness of the language environment. This information is summarized in the Results section.

More precisely, to collect information on the relative use of English/Arabic in the home, parents described, using a 1–5 scale, how much English and Arabic was used in the household by each relative (1 = Mainly or only Arabic (English: 0–20%, Arabic: 80–100%), 2 = Usually Arabic/English sometimes (English: 30%, Arabic: 70%), 3 = Arabic and English (English: 50%, Arabic: 50%), 4 = Usually English/Arabic sometimes (English: 70%, Arabic: 30%), 5 = Mainly or only English (English: 80–100%, Arabic: 0–20%)). While this information was collected for each member of the family in terms of output given to and received from the child, we calculated composite scores of relative Arabic/English use with parents, on the one hand, and siblings, on the other, with higher numbers indicating more English use.

To estimate the frequency of English and Arabic oral language and literacy activities (language environment richness), parents were asked to indicate, for each language, how many hours per week participants spent (1) reading and writing (including books, messaging, and homework), (2) speaking and listening (such as TV, music, and Skype), (3) taking part in extra-curricular activities (e.g., sports, clubs, and religious services), and (4) playing with friends. The measure of Arabic richness also included the number of hours per week of heritage-L1 classes. Parents estimated the frequency of these activities using a 1–5 scale: (1 = 0–1 h, 2 = 1–5 h, 3 = 5–10 h, 4 = 10–20 h, 5 = 20+ h). From these scales, a richness proportion score (0–1) was calculated for each language, where 1 would indicate high frequency of language-rich activities.

2.2.2. Matrix Analogies Test (MAT)

Participants completed two subtests of the Matrix Analogies Test (MAT; Naglieri 1985): reasoning by analogy and spatial visualization. These subtests ask participants to choose the picture that best completes a matrix. We used this test to measure non-verbal analytical reasoning. A raw compound score was obtained by adding the scores of the two subtests. The Cronbach's α coefficient of reliability for the compound scores was 0.87. This task was administered once, at T1, since it measures inherent cognitive abilities unlikely to change substantially within one year.

2.2.3. Quasi-Universal Nonword Repetition Task

We used Chiat's LITMUS nonword repetition task as a measure of phonological (i.e., verbal) short-term memory (Chiat 2015). This task asks participants to repeat 16 nonsense words that increase in length between two and five syllables. The 16 items include CVCV sequences only, with limited consonants, and conform to constraints on lexical phonology operative in many different languages (Chiat 2015). As such, performance on this test does not rely as much on English-specific phonotactics as other nonword repetition tests. Put differently, no non-word repetition test is completely language-independent; this version simply reduces the level of difficulty that might arise with English-specific phonotactics and was designed for use with bilingual children. Verbal instructions for this test are limited and were given in English. The Cronbach's α coefficient of reliability for this test was 0.64.

Participants carried out this task with noise-cancelling headphones. The research assistant played the recording of each item and participants were only allowed to listen to each item once. Participant repetitions were recorded for posterior scoring. A whole-item accuracy score was produced. That is, each of the 16 items was scored as either correct (1) or not (0) to produce a total score out of 16 (see Boerma et al. 2015 on alternative scoring schemes for this task). Repetitions that included omissions or replacements were scored as

incorrect. Allowances were made for target words that included a /p/ phoneme, which does not exist in Syrian Arabic. As such, repetitions were scored as correct when the only substitution for /p/ was by its voiced counterpart /b/.

Like the MAT, this test was only administered once because it measures inherent cognitive abilities. Nevertheless, this test was administered at T2 because, even though we used a quasi-universal task, some language-specific abilities are always implicated in non-word repetition tasks. Administering this task at T2 ensured that participants had more exposure to their English-L2, and thus were more bilingual. In other words, administering this task at T2 was intended to ensure a more reliable estimate of short-term verbal memory abilities.

2.2.4. Sentence Repetition Task (SRT)

An experimental SRT was used to gather information on participants' morphosyntactic abilities in the two languages. In SRTs, participants listen to a sentence and are asked to repeat it as closely as possible to the original one. Research has demonstrated that SRTs effectively measure language ability at the lexical, speech production, and morphosyntactic level (Klem et al. 2015; Polišenská et al. 2015). Therefore, even though this task implicates short term memory skills, at the same time it primarily implicates linguistic skills (Andreou et al. 2020; Marinis and Armon-Lotem 2015). We chose sentence-recall ability as our measure of language abilities for two reasons. First, an SRT was chosen because there is a strong link between sentence repetition and morphosyntactic competence, attesting to construct validity (Marinis and Armon-Lotem 2015; Polišenská et al. 2015). Second, the task does not make heavy demands on metacognitive and metalinguistic skills and is easy to administer to children in their homes or at school. We reasoned that a more highly demanding task might underestimate these children's language abilities because they were new arrivals who were undergoing acculturation that for most included adaptations to school culture because they had interrupted or no schooling prior to arrival in Canada (Paradis et al. 2020a; Paradis et al. 2021).

The SRT used in this study had 32 stimuli (one practice item plus 31 scored items). The stimuli for the SRT were based on Marinis and Armon-Lotem's (2015) LITMUS SRT for English. In the LITMUS SRT, sentences range in syntactic complexity, and thus, in difficulty (Marinis and Armon-Lotem 2015). The English version consisted of the morphosyntactic structures listed in Table 1. The English SRT stimuli were recorded by a native speaker of Canadian English.

Table 1. Structures used in the English sentence repetition task (SRT).

Name	Syntactic Structure
Declarative (*k* = 6)	Simple, active declarative sentences with auxiliaries and/or modals
Short passive (*k* = 3)	Passive sentences without an agent phrase
Long passive (*k* = 3)	Passive sentences with an agent phrase
Question (*k* = 6)	Object questions
Coordinated (*k* = 3)	Coordinated sentences
Subordinate (*k* = 4)	Conditional/temporal subordinate clauses
Relative (*k* = 6)	Object relative clauses (Object-Object and Subject-Object)

The Syrian Arabic SRT was developed using the English SRT as a model to ensure comparability of the two tasks. Syrian Arabic was chosen over Modern Standard Arabic (MSA) for this task for two reasons. First, Syrian Arabic is the variety children use in the home. Second, using MSA would have disadvantaged the participants who had no or little experience with schooling in Arabic, since schooling is the foundation for developing MSA skills. The Arabic SRT stimuli were recorded by an adult native speaker of Syrian Arabic, who was a recently arrived refugee. The morphosyntactic structures on the Arabic task are described in Table 2.

Table 2. Structures used in the Arabic SRT.

Name	Syntactic Structure
Declarative ($k = 6$)	Simple active declarative sentence with modals and/or particles
Short passive ($k = 3$)	Short passive (without an agent phrase)
Topicalization ($k = 3$)	Topicalization
Question ($k = 6$)	Object questions
Coordinated ($k = 3$)	Coordinated sentences
Subordinate ($k = 4$)	Conditional/temporal subordinate clauses
Relative ($k = 6$)	Object relative clauses (Object-Object and Subject-Object)

The English SRT and the Arabic SRT had a similar average number of morphemes per sentence (10.03 and 10.06, respectively). Comparing Tables 1 and 2 shows that there are comparable structures across the English and Arabic tasks, with the exception of long passives, which do not exist in Arabic and were replaced with topicalizations. Like long passives, topicalizations reverse the canonical order of agent and patient arguments and, in this sense, they are comparable in pragmatic function and complexity (on the functional equivalence between object topicalizations and long passives, see (Cowell 1964)). A complete list of stimuli for both the English and Arabic tasks are presented in Appendix A (Tables A2 and A1). Due to the low number of items per individual structure, this study does not investigate development on each particular structure but on the task as a whole. However, for the purposes of research question 2, structure was entered in the regression model to control for the variance that may be introduced due to some structures being more complex than others (see Section 4.3). In addition, Appendix B shows participants' performance on each individual structure at T1 and T2 for English and Arabic (Figures A1 and A2, respectively) for interested readers.

Administration of the SRT

Participants were presented with the 32 pre-recorded stimuli of the SRT, one at a time, using a laptop (PowerPoint) while wearing a noise-cancelling headset. There were two breaks during the task. Participants' repetitions were recorded for later transcription, scoring, and analysis. The first sentence, which was not scored, was a practice item. Participants were allowed to listen to the practice item more than once to ensure they understood the task mechanics but could only listen to the test items once.

Scoring

Performance on an SRT can be scored in different ways depending on the goal. For this study, we produced a score of morphosyntactic accuracy. This score measured whether participants had repeated the target morphosyntactic structure accurately or not regardless of other (lexical or morphological) errors in their repetition. This type of scoring narrowed down the aspects of performance to be assessed by disregarding errors that did not affect the target morphosyntactic structure. As such, this scoring was aimed at being a more direct proxy of morphosyntactic abilities than other types of scoring, e.g., ones that include all errors.

Productions that did not contain at least one inflected verb were considered unscorable. An example of an unscorable repetition for the target sentence *The team that my brother cheer-ed for won the race* appears in (1), and one example of an unscorable repetition for *l-ħsˤaːn əlli l-fællæħ dˤarab rafæs-ɔ ʕæ-dˤahr-ɔ* 'The horse that the farmer hit kicked him in the back', appears in (2). Unscorable responses were not numerous: at T1, 7.53% of responses in English and 2.66% of responses in Arabic and were unscorable. At T2, 1.11% and 1.23%, respectively, were unscorable. Since unscorable sentences could not be scored in terms of the correct/incorrect repetition of the target structure, these productions are not considered in the statistical analyses.

(1) The team for the race.
(2) l-ħsˤa:n   l-faras
    the-horse   the-horse

The requirements for each structure to be considered correct in participants' productions depended on the target structure. For example, for short passive structures to be scored as correct in English, participants' repetitions had to contain passive auxiliary *be* and the past participle form of the verb. Example (3) shows a participant's repetition of the target short passive stimulus "He was push-ed hard against the ground". This repetition was scored as incorrect because the morphology in the verb indicated that the passive structure had not been repeated accurately. In Arabic, participants' responses had to include the passive morphology in the form of a prefix on the verb. Example (4) shows a target passive sentence in Arabic and (5) shows a participants' response. This response was scored as correct even though the participant made other errors (i.e., they eliminated the article and the preposition). Regardless of these errors, the verbal prefix indicating passive voice was repeated correctly. To provide another example, for sentences that contained a subordinated clause to be scored as correct, participants' answers had to contain two clauses and a subordinator. Example (6) provides a repetition of the target sentence "The child ate breakfast after he wash-ed his face" in English, which was scored as correct despite some lexical and functional substitutions and additions. Example (7) shows a target sentence for the subordinate structure in Arabic and (8) a repetition that was scored as incorrect due to the substitution of the subordinator *iza* by the coordinator *wu*.

(3) He was push-*ing* hard against the ground.
(4) n-dæfæʃ   b-ʔuwe   ʕæ-l-ʔardˤ
    PASS-pushed   in-hard   to-the-ground
    'He was pushed hard against the ground.'
(5) n-dæfæʃ   ʔuwe   ʕæ-ʔardˤ
    PASS-pushed   hard   to-ground
    'He was pushed hard against the ground.'
(6) *My children* ate *the* breakfast after he wash-ed his face.
(7) rɒħ   y-a:xd-u   l-wlæ:d   hdi:ə
    FUT   3-get-PL   the-children   gift
    izæ   nadˤdˤaf-u   l-be:t
    if   clean-3PL   the-house
    'The children will get a present if they clean the house.'
(8) y-a:xd-u   l-wlæ:d   hdi:ə
    3-get-PL   the-children   gift
    wu   nadˤdˤaf-u   l-be:t
    and   clean-3PL   the-house
    'The children get a present and they clean the house.'

In order to ensure close control and full reliability of this scoring, every sentence was scored independently twice by researchers with syntactic training and all disagreements were settled by group discussion with two of the co-authors.

2.3. Data Analysis Procedures

All statistical descriptive and inferential analyses were performed in R (version 4.0.3; R Core Team 2018). We discuss details in the modeling procedures and the R packages used for all of our analyses here. Complete model details are presented in tables in Appendices B and D and summary findings are presented in the Results section.

To address our first research question, we fit two logistic linear mixed-effects regression models using the package *lme4* (version 1.1–25; Bates et al. 2015), predicting whether a morphosyntactic structure of a given item would be repeated accurately (1) or not (0). In the first of these two models, the predictors (i.e., the fixed effects) were Language (a factor with two levels: Arabic and English), Time (a factor with two levels: T1 and T2), and the interaction between Language and Time. The random effect structure included one intercept for Item and one intercept for Participant. The random effects also included two

by-Participant varying slopes: for Language and Time. The random slopes were considered relevant due to the fact that the effect of Language and Time may vary across Participants. While the interaction between Language and Time could have been a reasonable by-Participant slope, it could not be estimated by the model. The second model specifically investigated the role of AOA in accuracy in the two languages and the two time points. This model was the same as the one described above, with the only difference being the inclusion of an additional fixed effect, AOA group, as well as the corresponding three-way interaction between Language, Time, and AOA group. Visualizations of the predicted probabilities derived from these models were obtained with the package *sjPlot* (version 2.8.7; Lüdecke 2021).

To address our second research question, we also employed logistic linear mixed-effects regression using the package *lme4* (Bates et al. 2015). Because our second research question concerns associations between concurrent individual difference factors and SRT performance, we generated four models, one for each language and each time interval. In all the models, whether the morphosyntactic structure of a sentence had been repeated correctly or not was set as the outcome variable (accurate repetitions were given a 1, and incorrect repetitions were given a 0). The fixed effects for T1 models included English AOA, length of English or Arabic schooling, English or Arabic richness, years of maternal education, language use with siblings, non-verbal analytical skills (raw scores out of 32), nonword repetition (raw scores out of 16), and morphosyntactic structure type of the stimuli. All fixed effects were numerical except for morphosyntactic structure, which was a factor (6 levels/structures in English and 7 levels/structures in Arabic; short and long passives were combined for English). Numerical predictors were centered around 0 and standardized using the *scale* function from the base package. This procedure ensures that the slopes of the fixed effects are comparable and interpretable in relation to one another (Winter 2019). The fixed effects for the T2 models included the same factors as the T1 models but concurrent at T2. Predictors that remained stable between T1 and T2 (e.g., English AOA, maternal education) were entered as such.

For all models, we conducted backwards selection of the fixed effects. That is, one by one, non-significant predictors were eliminated from the model to arrive at the most parsimonious model. At each step, the simpler model was compared to the fuller model in terms of Akaike information criterion (AIC) values and employing log-likelihood ratio tests. The simpler model was retained if it did not entail a significant loss of goodness-of-fit. However, certain fixed effects were retained in all models regardless of significance. Morphosyntactic structure was retained as a language-level factor because the SRT varied in complexity of structures, and retaining this factor accounted for variance associated with complexity. Similarly, the cognitive measures, non-verbal analytical reasoning and nonword repetition, were retained in each model, as a child-level factor, because these were raw scores rather than age-controlled, standardized scores, and our sample had an age range. Finally, the random effect structure for all models included one intercept for Participant and one for Item. This was the most complex random-effect structure that the models could support. Post-hoc comparisons on the morphosyntactic structure factor were carried out using the package *emmeans* (version 1.5.4; Lenth 2021).

To estimate the goodness-of-fit of the models, the C-index of concordance was calculated using the *Hmisc* package (version 4.4-1; Harrell 2020; to know more about the C-index and how to compute it see Baayen 2008). Because we used logistic regression, R^2 values could not be calculated, and so C-index was employed. Models with a C-index of 0.80 or above are considered to have a good fit (Hosmer et al. 2013). All models were above this threshold. Finally, data visualizations were created using *ggplot2* (version 3.3.2; Wickham 2016).

3. Results

Information on the characteristics of the participants and their families at T1 and T2 are displayed in Table 3. Since the sample at T1 and T2 remained the same, variables that did not change over time are only presented once.

Participants' mean AOA was around 7–8, indicating that children were sequential bilinguals with prolonged heritage-L1 exposure prior to the onset of bilingualism. Despite this exposure, at T1 the length of participants' schooling in Arabic was significantly shorter than their schooling in English, as indicated by a paired-samples t-test, $t(101) = 2.974$, $p = 0.004$. This result is indicative of interrupted schooling.

Most participant families had low socioeconomic backgrounds, as indexed by parental education levels. Most parents in the sample had only completed primary or secondary school, although there were some with post-secondary education. In addition, families were large, with an average of more than 4 children per household. During the period of this study, most families were living on social assistance.

Table 3. Participant and family characteristics at T1 and T2.

	T1	T2
Age (years)	9.51 (1.89)	10.40 (1.86)
Length of residency (months)	24.04 (7.20)	34.69 (7.42)
English AOA (years)	7.66 (1.94)	N/A
Length of English schooling (months)	19.40 (6.23)	27.92 (6.13)
Length of Arabic schooling (months)	14.26 (15.28)	N/A
Maternal education (years)	9.71 (4.06)	N/A
Paternal education (years)	9.89 (3.93)	N/A
Number of children in the family	4.44 (1.49)	4.67 (1.37)
Language use with parents (1–5 scale) [a]	1.10 (0.27)	1.21 (0.45)
Language use with siblings (1–5 scale) [a]	1.58 (0.96)	1.72 (1.10)
English richness [b]	0.44 (0.13)	0.50 (0.12)
Arabic richness [b]	0.41 (0.11)	0.40 (0.11)
Non-verbal reasoning (out of 32)	8.61 (5.89)	N/A
Verbal memory (out of 16)	N/A	11.78 (2.47)

Note: [a] 1 = Mainly or only Arabic (English: 0–20%, Arabic: 80–100%), 2 = Usually Arabic/English sometimes (English: 30%, Arabic: 70%), 3 = Arabic and English (English: 50%, Arabic: 50%), 4 = Usually English/Arabic sometimes (English: 70%, Arabic: 30%), 5 = Mainly or only English (English: 80–100%, Arabic: 0–20%). [b] Scores between 0 and 1 capturing frequency of language-rich activities, where values closer to 1 indicate higher frequency (1 = 0–1 h, 2 = 1–5 h, 3 = 5–10 h, 4 = 10–20 h, 5 = 20+ h).

Changes in children's language environments over time are noteworthy. First, language use with parents and with siblings was captured as a score between 1–5 (see Table 3 notes for correspondence between scores and relative proportional use of each language), with higher numbers indicating more English use. English was used more often with siblings than with parents at both T1 and T2 (T1: $t(101) = 5.810$, $p < 0.001$; Cohen's $d = 0.54$ (medium effect size); T2: $t(101) = 5.988$, $p < 0.001$. Cohen's $d = 0.48$ (small effect size)). Nevertheless, there was a significant increase in the use of English between participants and their parents from T1 to T2, $t(101) = 3.043$, $p = 0.003$; whereas, the proportion of English use with siblings remained largely stable, $t(101) = 1.3664$, $p = 0.175$. Second, for English richness, there was a significant increase at T2, $t(101) = 3.7713$, $p < 0.001$ (see Table 3 notes for correspondence between scores and frequencies of language-rich activities in an average week). This indicates that participants engaged more frequently in English-rich activities at T2. On the other hand, the richness of the Arabic environment stayed stable, $t(101) = 0.5536$, $p = 0.581$. Finally, regarding cognitive factors, note that the mean score for non-verbal analytic reasoning was low because these were raw scores (not age-corrected). However, the large variance across individuals on this task indicates its suitability as a co-variate in the modelling.

3.1. Changes over Time in Morphosyntax in Both Languages

To address our first research question regarding dominant language shift, we examined performance on the SRT in each language and over time. The distribution of participants' scores in both languages and at both times appears in Figure 1, where the maximum score is 31. Mean scores and SDs for each language and time appear in Table A3

in Appendix B. Figure 1 suggests that, at both T1 and T2, children's scores tended to be higher and less variable in Arabic than in English, but with some overlap between the two languages. As mentioned above, in this study we did not focus on performance with each individual structure, but we include the descriptive results in Appendix C.
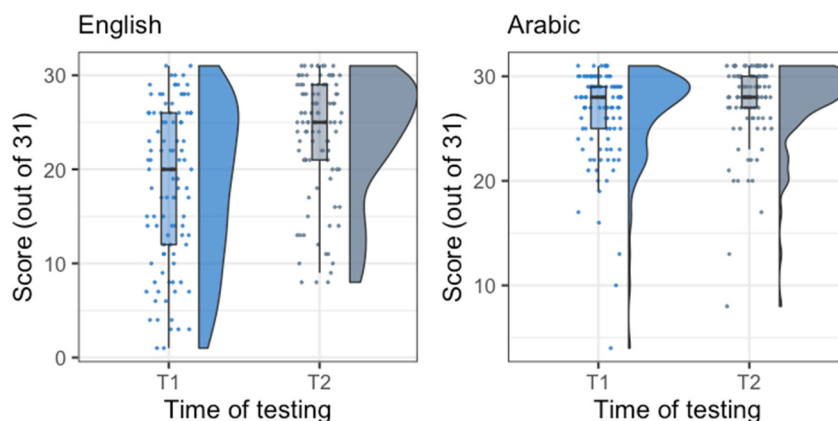

Figure 1. Participants' morphosyntactic accuracy score at T1 and T2 in English (**left**) and Arabic (**right**). Each point is one participant, and it shows their total score out of 31 (on the *y*-axis). Points appear jittered horizontally due to discreteness (i.e., point overlap). The split violin, to the right of the points, is a density plot indicating the distribution of the data. Peaks in the density plot indicate where participants are concentrated.

Participants' scores at T1 and T2 were correlated positively and strongly for both languages (English: *r*(100) = 0.810, *p* < 0.001; Arabic: *r*(100) = 0.806, *p* < 0.001). As described in the data analysis procedures (Section 2.3), we fit a logistic mixed-effects regression in order to determine the trends in development in the two languages for the whole sample. In this model, participants' morphosyntactic accuracy score (either 1 or 0) was predicted by Language (English/Arabic), Time (T1/T2), and the interaction between Language and Time. The full output of this model appears in Tables A4 and A5 in Appendix B. Considering that Language, Time, and the interaction between the two factors were all significant, the overall findings of this model are as follows: Participants were more accurate on the task in Arabic than in English. Accuracy improved in the two languages between T1 and T2, but the improvement was significantly larger in English, indicating a narrowing of the between-language gap at T2. To facilitate interpretation of these results, the predicted probabilities of a correct morphosyntactic structure repetition according to Language and Time, based on the model described here, appear in Figure 2.
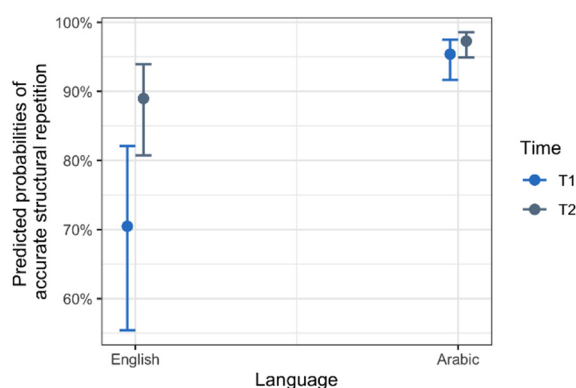

Figure 2. Marginal effects of the Language*Time interaction in the logistic mixed-effects model comparing performance in the two languages at the two times. The point indicates the predicted average probability of a morphosyntactic structure being repeated accurately according to Language and Time. The whiskers indicate the 95% confidence interval.

Changes over Time according to Age of L2 Acquisition Onset

As outlined in the Introduction, dominant language shift may depend partly on the age of first exposure to the L2, so that the children in our sample with earlier AOAs may be more likely to shift dominance to the L2 than those with older AOAs from T1 to T2. To investigate if there was a difference in function of AOA, we split participants into two groups according to whether they fell above or below the average English AOA. This division thus yielded two groups of similar size ($N_{younger}$ = 50; N_{older} = 52) with different AOAs ($M_{younger}$ = 6.00, $SD_{younger}$ = 0.95; M_{older} = 9.25, SD_{older} = 1.14). Figure 3 presents the performance of these two subgroups at both times for English and Arabic. Mean scores and SDs for each subgroup, for each language and time are included in Table A3 in Appendix B.
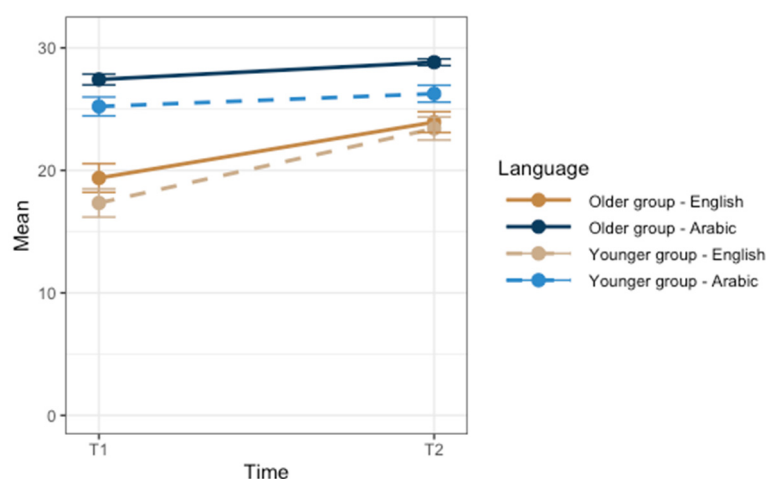

Figure 3. Participant performance at T1 and T2 in English (brown) and Arabic (blue) according to age at L2 acquisition onset (AOA) group.

To assess whether language shift would occur as a function of AOA, and as described in the data analysis procedures in Section 2.3, we fit a logistic mixed-effects regression model with a triple interaction between Language, Time, and AOA group (a factor with two levels: Older vs. Younger). The results for this model can be seen in Tables A6 and A7 in Appendix B. Focusing on the results pertinent to AOA, the overall findings of this model are as follows: The Younger subgroup had significantly lower scores in Arabic at T1 than the Older subgroup; moreover, the Younger subgroup made significantly smaller gains in Arabic between T1 and T2 than did the Older subgroup. On the other hand, the Younger subgroup made larger gains in English than the Older subgroup over time. Therefore, the difference in performance between English and Arabic had shrunk significantly more between T1 and T2 for the Younger subgroup than for the Older subgroup. To facilitate interpretation of these results, the predicted probabilities of a correct morphosyntactic structure repetition according to Language, Time, and AOA subgroup, based on the model described here, appear in Figure 4.
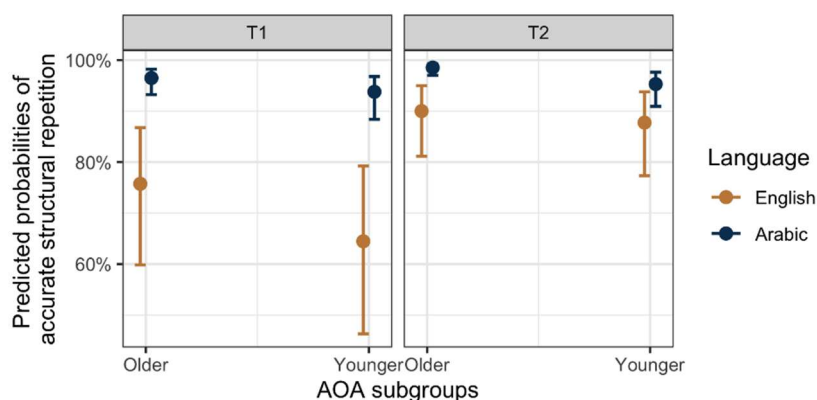
Figure 4. Marginal effects of the Language*Time*AOA interaction in the logistic mixed-effects model comparing performance in the two languages at the two times by the two AOA subgroups. The point indicates the predicted average probability of a morphosyntactic structure being repeated accurately according to Language, Time, and AOA subgroup. The whiskers indicate the 95% confidence interval.

3.2. T1: Influence of Cognitive, Age, and Input Factors on Performance in English-L2 and Syrian Arabic-L1

In this section, we report the results of our modeling analyses aimed at addressing our second research question. We only present the significant results and visualizations in this section. The full output of the most parsimonious models appears in Appendix D (Tables A8–A12). Significant factors are reported according to the size of their contribution to explaining variance in children's performance, shown in the Beta coefficient values as these factors were centered and standardized.

In English at T1, maternal education made the largest contribution to the model, and in the expected direction, where children of higher educated mothers were more accurate on the SRT ($\beta = 0.558$, SE = 0.138, z-value = 4.055). More English use with siblings was also associated with greater accuracy on the SRT ($\beta = 0.534$, SE = 0.151, z-value = 3.540) and made a similar contribution as maternal education to the variance in performance. Next, children with stronger verbal memory skills ($\beta = 0.376$, SE = 0.140, z-value = 2.689) and stronger non-verbal analytical skills ($\beta = 0.368$, SE = 0.149, z-value = 2.466) performed better than children with weaker skills. Children schooled for longer in Canada ($\beta = 0.361$, SE = 0.136, z-value = 2.657) had better performance on the SRT. Similarly, children with an older AOA ($\beta = 0.300$, SE = 0.150, z-value = 2.005) and higher English richness ($\beta = 0.274$, SE = 0.139, z-value = 1.969) were more accurate. The effect of each significant predictor can be visualized in Figure 5. The C-index for the model predicting performance in English at T1 was 0.89. It should be noted that morphosyntactic structure was not significant in this model; however, it was left in the model to control for any variance associated with it.

The optimal model for Arabic at T1 indicated that verbal memory ($\beta = 0.522$, *SE* = 0.127, z-value = 4.099) made the largest contribution to accounting for variance in children's SRT scores. Next, older English AOA was also significantly associated with better performance on the SRT ($\beta = 0.325$, *SE* = 0.133, z-value = 2.452). Finally, complexity of morphosyntactic structure influenced accuracy in performance. A post-hoc test with a Tukey adjustment (Appendix D, Table A12), found that relative structures were the least likely to be repeated correctly compared with other structures, and topicalizations were less likely to be repeated than passives. The effect of each factor can be visualized in Figure 6. The C-index for this model was 0.91.
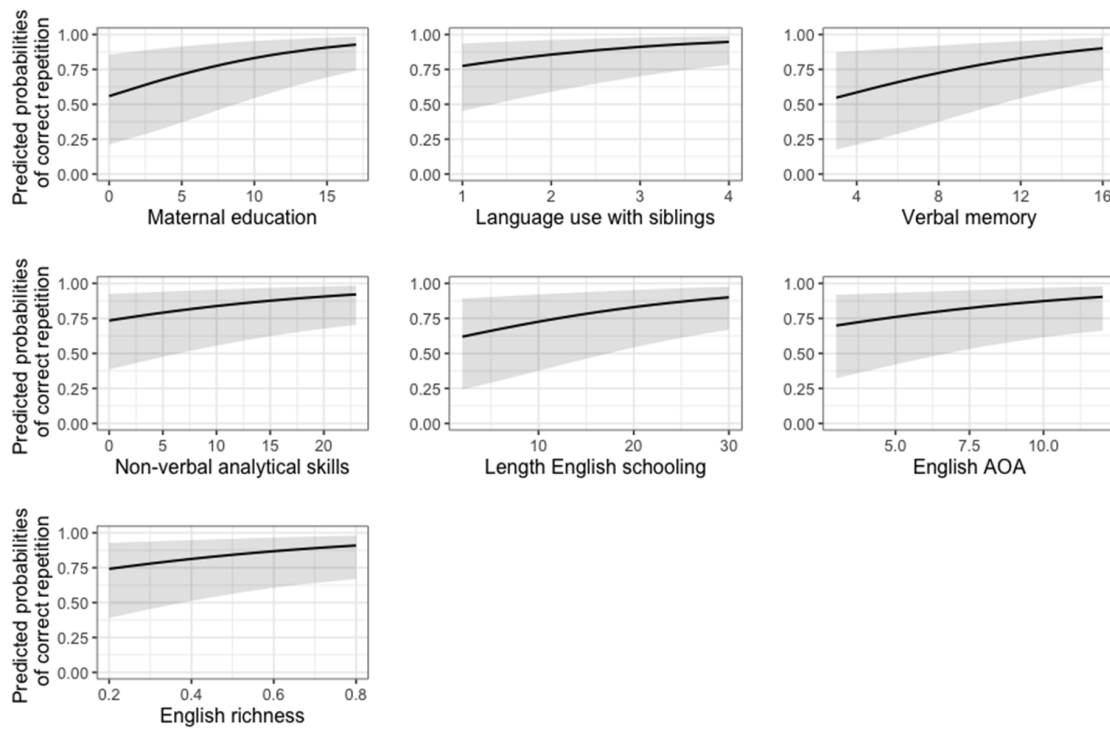
Figure 5. Predicted probabilities of a correct response given each fixed effect in the optimal model of English T1, predicting correct repetition of target morphosyntactic structure.
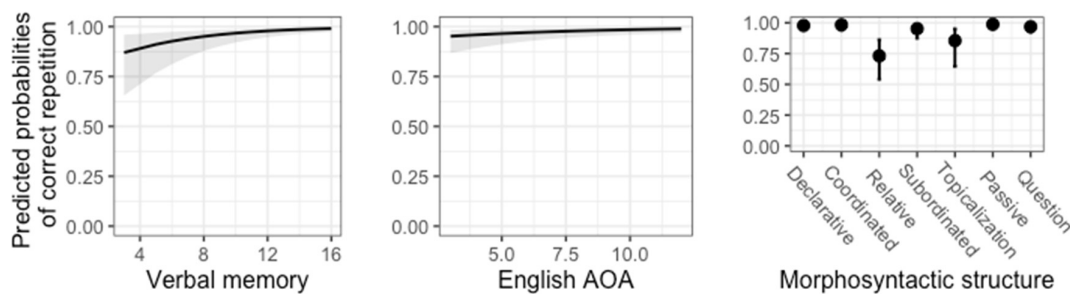
Figure 6. Predicted probabilities of a correct response given each fixed effect in the optimal model of Arabic T1, predicting correct repetition of target morphosyntactic structure.

3.3. T2: Influence of Cognitive, Age, and Input Factors on Performance in English-L2 and Syrian Arabic-L1

As with the preceding section, we only present the significant results and visualizations here. The full outputs of the most parsimonious models are in Appendix D Tables (A13–A18). Significant factors are reported according to the size of their contribution to explaining variance in children's performance, shown in the Beta coefficient values as these factors were centered and standardized.

At T2 in English-L2, the factor that made the largest contribution to the optimal model was verbal memory ($\beta = 0.758$, $SE = 0.153$, z-value = 4.968), followed by non-verbal analytical skills ($\beta = 0.553$, $SE = 0.157$, z-value = 3.532); therefore, children with stronger cognitive skills were more accurate on the SRT at T2. As with T1, there was an association between more English use with siblings and better performance ($\beta = 0.392$, $SE = 0.153$, z-value = 2.559). Differently from the optimal model for T1, morphosyntactic structure contributed to variance in performance in English at T2. The post-hoc test with a Tukey adjustment (Appendix D, Table A15), showed that relative clause structures were less likely to be repeated correctly than coordinated and subordinated clauses; there was also a trend for passives to be less accurately reproduced compared to coordinated clauses. While the

effect of maternal education was not significant at T2, and thus, not present in the optimal model, it showed a trend toward significance (*p* = 0.054). The effect of each predictor can be visualized in Figure 7. The C-index for this model was 0.91.

Figure 7. Predicted probabilities of a correct response given each fixed effect in the optimal model of English T2, predicting correct repetition of target morphosyntactic structure.
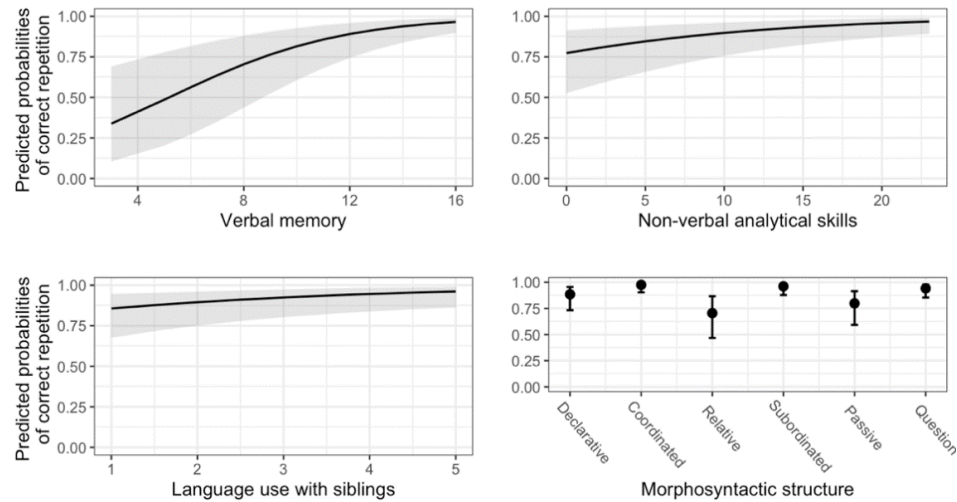
In Arabic at T2, older English AOA made the largest contribution to the model and in the expected direction: children first exposed to English at older ages performed better (*β* = 0.619, *SE* = 0.143, *z*-value = 4.317). The factor with the second largest contribution was language use with siblings (*β* = −0.401, SE = 0.125, *z*-value = −3.203). This was a negative predictor, indicating that more English use with siblings was associated with lower accuracy on the task. In terms of cognitive skills, verbal memory (*β* = 0.387, *SE* = 0.126, *z*-value = 3.075) made a larger contribution to the model than non-verbal cognitive skills (*β* = 0.249, *SE* = 0.145, *z*-value = 1.718), and the latter showed only a trend towards significance. Finally, morphosyntactic structure was a significant predictor at T2 as it was at T1. The post-hoc contrasts (Appendix D, Table A18) showed that relative clause structures were significantly less likely to be repeated accurately than all other structures, except for topicalizations. The effect of each factor can be visualized in Figure 8. This model had a C-index of 0.92.
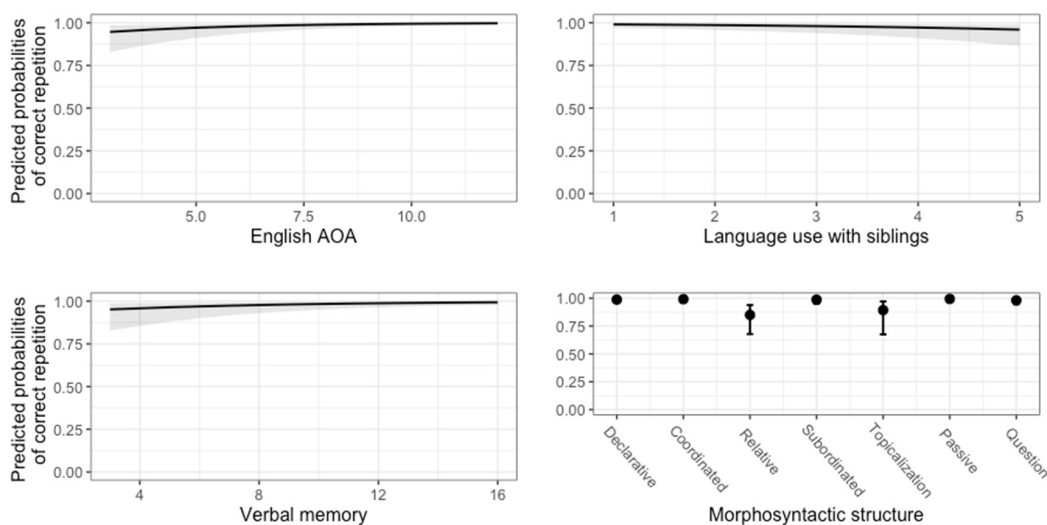

Figure 8. Predicted probabilities of a correct response given each fixed effect in the optimal model of Arabic T2, predicting correct repetition of target morphosyntactic structure.

4. Discussion

This longitudinal study examined morphosyntactic development in the Arabic heritage-L1 and English-L2 of first-generation Syrian refugee children in Canada. Participants had resided in Canada for 2 years at T1 and 3 years at T2, and ranged in age from 6–13 at T1. Our objectives were (1) to determine if dominant language shift from the L1 to the L2 was taking place early in their residency in the host country, and (2) to assess the role of cognitive, age, and language environment (input) factors as sources of individual differences in their morphosyntactic development in each language and time interval. In so doing, we sought to contribute uniquely to the research on bilingual morphosyntactic development by adopting a longitudinal design, including first-generation migrants, and examining outcomes in both languages and the factors influencing individual differences in those languages.

4.1. Dominant Language Shift

After three years of residency in Canada, the morphosyntactic skills in both the English-L2 and Arabic heritage-L1 of these Syrian refugee children were still developing as they increased from T1 to T2 in both languages; however, growth was steeper in the L2 (Figure 2). The steeper growth in English is possibly due to the high-quality exposure through L2 schooling; although, children's higher scores on the task in Arabic would mean there was less room for growth in this language, as the task is bounded (there is a maximum score). Notably, even at T2, performance was still higher in Arabic than English. Regarding the role of AOA in dominant language shift, both younger AOA and older AOA groups had superior Arabic performance at T1 and T2 than English performance (Figures 3 and 4). The younger group displayed flatter growth between T1 and T2 in Arabic than the older group, suggesting a hint of stagnation in heritage-L1 morphosyntactic development that was not evident in the older group. In sum, dominant language shift to English-L2 had not yet taken place for the group as a whole nor for the younger and older subgroups (but see AOA considered as an individual difference factor below).

Contrary to expectations based on refugee background, the children in this study did not seem to be at greater risk of rapid L1 attrition, in spite of absence of age-appropriate levels of schooling in their L1 (see Table 3). Thus, our participants did not shift to L2 dominance after 3 years of exposure to the L2. These findings appear to contrast with those of Jia and Aaronson (2003), whose data showed dominant language shift in preference, use, and proficiency for younger AOA participants during the first 3 years of residency. However, parent report on heritage-L1 attrition showed greater loss of reading and writing skills than listening and speaking skills in this study. Our results seem to fall in line, however, with those of Hamann et al. (2020) who found strong SRT performance in Arabic for first-generation Arabic heritage-L1 children with a similar age range and time of residency in Germany. Compared with G. Jia et al. (2014) and Pham and Kohnert (2014), this study highlights differences in heritage-L1 dominance between school-age first-generation versus second-generation heritage language children.

4.2. Sources of Individual Differences in Children's Morphosyntactic Development in English and Arabic

Overall, children were more accurate on the SRT in Arabic at both time intervals, and this reduced variance in performance is a possible explanation for fewer significant fixed effects in the models at T1 and T2 than for English. Nevertheless, our results revealed similarities and differences in how factors influenced each language, as well as some changes over time.

4.2.1. Stable Effects of Cognitive Abilities

Cognitive abilities, as measured by verbal memory and non-verbal analytic tasks, were associated with morphosyntactic abilities in both languages and at both time intervals. Verbal memory was more closely associated with SRT performance than non-verbal skills

in both languages, which was expected given SRT does involve working memory capacity (although, collinearity was not an issue because SRT scores and verbal memory scores were only mildly correlated, $r = 0.30$, $p = 0.002$). Of particular note is that these cognitive abilities often accounted for more variance than age and input factors. For example, individual differences in cognitive abilities made a larger contribution to variance than AOA at T1 English and T1 Arabic, and verbal memory predicted more than input factors at T2 English. Taken together with studies discussed in the introduction, we can observe that verbal memory and non-verbal cognitive abilities play a role in lexical, morphological, and morphosyntactic acquisition in both the heritage-L1 and majority-L2 of bilingual children (Hamann et al. 2020; Paradis et al. 2020a; Pham and Tipton 2018).

4.2.2. Differential Effects of Age at L2 Acquisition Onset

We found that older AOA was advantageous for children's Arabic-L1 at both time intervals but was only advantageous to their English-L2 at T1. Thus, findings for Arabic are in line with many existing studies showing that a delay in exposure to the societal L2 promotes heritage language development and maintenance (Albirini 2018; see Montrul 2016, for review). However, our findings contrast with those of Hamann et al. (2020), who found AOA to not be associated with SRT performance in Arabic among Arabic-German bilinguals. This difference could be due to sample size; the present study had a much larger sample, or due to sociocultural context factors. Future, cross-national comparative research would be needed to understand more.

For children's English-L2 morphosyntactic development, those who arrived when they were older tended to have higher scores on the SRT, and this is consistent with other studies at the early stages of L2 development (e.g., Chondrogianni and Marinis 2011; Paradis 2011). We found no evidence for an older AOA advantage at T2, however, which suggests that the older AOA advantage might recede with increased L2 exposure (cf. Jia and Aaronson 2003; Jia and Fuse 2007). However, caution is needed in interpreting null effects.

4.2.3. Differential Effects of Language Environment

Overall, participants lived in largely Arabic-dominant homes at T1 and T2 (see Table 3). Nevertheless, there was a trend in the direction of using more English in the home. The use of English-L2 in sibling interactions had a positive influence on morphosyntactic development in this language at T1 and T2, but showed a negative influence on the Arabic-L1 at T2. Because sibling interaction is a relative variable—more English means less Arabic—the differential and changing contribution to each language's development over time is in line with prior cross-sectional research on L2 (Hamann et al. 2020; Paradis et al. 2020a; Rojas et al. 2016; Sorenson Duncan and Paradis 2020b) and heritage-L1 development (Pham and Tipton 2018; Rojas et al. 2016; Sorenson Duncan and Paradis 2020b).

Similar to language use at home, maternal education and richness of the L1 and L2 environment had differential impacts on language and time period. Higher maternal education was the largest predictor of stronger English-L2 morphosyntax at T1, and greater richness in the English-L2 environment also contributed to stronger performance in English at T1. However, these input factors were not associated with performance in Arabic at either time interval and were only marginally predictive of individual differences in English at T2. This pattern of differential impact of maternal education on the majority-L2 versus the heritage-L1 has also been found in other studies (Armon-Lotem et al. 2011; Hamann et al. 2020; Place and Hoff 2016; Prevoo et al. 2014; Rojas et al. 2016). The relatively low impact of both these environment factors could be because newly arrived refugee families, living mainly on social assistance might not have the resources to provide a rich home environment in either language, even in families where mothers had some higher education (Paradis et al. 2020b). The shift in significant influence of maternal education and English richness at T1 to marginal or non-influence at T2 could signal a kind of "ceiling" effect for these factors in the L2 development of children from under-resourced families. Another explanation could be that the development of some linguistic domains, such as complex

morphosyntax, is less sensitive to variation in language richness and maternal education than some other domains. However, this explanation is partially undermined by several studies that have shown language richness and maternal education to have an impact on L2 morphology and syntax in the first three years of dual language learning (Jia and Aaronson 2003; Paradis 2011; Paradis et al. 2017; Paradis et al. 2020a). A third explanation could lie in the task used to measure morphosyntax or in our scoring system which focused on accuracy in repetition of the morphosyntactic structure over general repetition accuracy (cf. Hamann et al. 2020).

4.3. Morphosyntactic Complexity on the Sentence Repetition Task

Analyzing outcomes for individual structures on the SRT was not a focus of this study. Even though our task included structures of different complexity levels, there were too few exemplars of each structure to warrant systematic analysis. However, structure was included, as a language-level factor, in the models to specify variance in performance due to morphosyntactic complexity/difficulty of stimuli. Previous research has shown that more complex stimuli on an SRT should be associated with differences in performance (Chiat et al. 2013; Komeili et al. 2020). In this study, complexity differentially impacted the L1 and the L2. At T1, structure was not a significant fixed effect in the model for the English SRT, meaning children's performance in their L2 was not sensitive to greater complexity/difficulty in the stimuli. At T2, structure was a significant fixed effect in English. By contrast, structure played a role at both T1 and T2 in Arabic. In all cases when structure was a significant effect, one or more complex structures were associated with lower accuracy. Taken together, we interpret these findings to indicate that at T1, the children's English proficiency had not advanced sufficiently for them to be sensitive to morphosyntactic complexity on the SRT.

5. Conclusions

This study suggests that rapid heritage-L1 attrition and dominant language shift to the L2 is not evident in morphosyntax in the first 3 years for first-generation child migrants. Even though dominance in the L2 is generally expected for heritage-L1 children by adulthood, these first-generation children might have a good chance of strong heritage language maintenance in the long term. However, our results are based on participants who arrived as a large cohort—allowing for strong social connections within the community, and also arriving in large families—meaning more interlocutors at home. Therefore, the social context and language environment of the Syrian children in this study could have promoted heritage language use and maintenance more than would be the case for other first-generation children who might have fewer opportunities to hear and use the heritage language.

Regarding individual differences, this study revealed a strong role for cognitive abilities in shaping development in both languages and over time, in some cases explaining more variance than input and age factors. In contrast, we found that age and input factors could affect the heritage-L1 and majority-L2 of early bilinguals differentially, and that the effects of these factors can change over time. Notably, older AOA is a consistent and strong benefit to individual children's heritage-L1 development, but only shows a short-term benefit to the rate of their L2 development. Finally, L2 use at home among siblings can have positive short-term effects for the L2 but negative effects in the longer term for the heritage-L1. The last result raises the issue of tension between promoting the societal-L2 while protecting the heritage-L1 because sibling interaction is a malleable factor (unlike inherent cognitive skills or AOA). We believe that when advising families about language use at home, encouraging use of the L1 over the L2 among siblings should be the message. This is because many other factors outside the home can contribute to growth in L2 skills, while home language use is likely to be more important for promoting and protecting the heritage language.

Appendix A. List of Stimuli Used in SRT

Table A1. Stimuli used in English SRT.

Structure	Example
Declarative ($k = 6$)	She can bring the glass to the table. They have been rid-ing the bicycle around the backyard. They are eat-ing the banana-s in the park. The kitten could have bounc-ed the ball down the stair-s. The boy must sweep the floor in the kitchen. The teacher has been look-ing at us all day.
Short passives ($k = 3$)	She was stopp-ed at the big red light-s. The children were tak-en to the office. He was push-ed hard against the ground.
Long passives ($k = 3$)	The cow was kick-ed in the leg by the donkey. She was se-en by the doctor in the morning. The mother was follow-ed by the girl.
Question ($k = 6$)	What did the mother cook in the evening? Who have they se-en near the front door? Which picture did he paint at home yesterday? What did the father buy last month? Who did the girl meet in the library yesterday? Which drink did the neighbour spill in the house?
Coordinated ($k = 3$)	The mother is shopp-ing and the child is study-ing at home. The dog bark-s outside and the child crie-s inside. Our neighbor clean-s the car and his son play-s basketball.
Subordinate ($k = 4$)	If the weather is warm, we can go to the park. Before the girl eat-s dinner, she will play with the computer. The children will get a present if they clean the house. The child ate breakfast after he wash-ed his face.
Relative ($k = 6$)	The boy that the neighbor help-ed has lost his way. They should wash the baby that the mother is feed-ing. The horse that the farmer push-ed kick-ed him in the back. The mother made the meal that the children are eat-ing. The children enjoy-ed the candy that they tast-ed. The team that my brother cheer-ed for won the race.

Table A2. Stimuli used in Syrian Arabic SRT.

Structure	Example					
Declarative (*k* = 6)	ʕæm PROG	yæ-kl-u 3-eat-PL	l-mɔ:z the-bananas	bə-l-jne:ne in-the-park		
	'They are eating the bananas in the park.'					
	rfi:ʔ-i friend-my	byə-ʔder 3SG.M-can	yə-rfæʕ 3SG.M-lift	hæ-l-kersi this-the-chair	l-kbi:r the-big	
	'My friend can lift this big chair.'					
	læzem must	t-ħɒtˤ 3SG.F-put	l-kæ:se the-cup	ʕæ-tˤ-tˤawle on-the-table		
	'She must put the glass on the table.'					
	mæ NEG	bt-eħsen 3SG.F-can	l-ʔænse the-teacher	t-ji:b 3SG.F-bring	l-ktæ:b the-book	ʕæ-sˤ-sˤaf to-the-class
	'The teacher cannot bring the book to the class.'					
	mæ NEG	ʃereb drank	l-wælæd the-child	ħæli:b milk	sɒxən hot	əs-sɒbəħ the-morning
	'The child did not drink hot milk in the morning.'					
	mæ NEG	ʔære:-t read−1SG	ʔesˤa story	ħelwe beautiful	bə-l-mædræse in-the-school	
	'I did not read a beautiful story at school.'					
Short passive (*k* = 3)	ət-æxæd−u PASS-taken-3PL	l-wlæ:d the-children	ʕæ-l-mæktæb to-the-office			
	'The children were taken to the office.'					
	n-dæfæʃ PASS-pushed	b-ʔuwe in-hard	ʕæ-l-ʔardˤ to-the-ground			
	'He was pushed hard against the ground.'					
	t-xæ:laf-et PASS-fined-3SG.F	ʕæ-l-iʃæ:ra on-the-light	l-ħæmra the-red			
	'She got a fine at the red light.'					
Topicalizations (*k* = 3)	l-ʔm the-mother	leħʔ-æ followed-her	esˤ-sˤabi the-boy	ʕæ-ʃ-ʃa:reʕ to-the-street		
	'As for the mother, the boy followed her to the street.'					
	l-baʔara, the-cow	l-ħma:r the-donkey	dˤarab-æ hit-her	barra outside		
	'As for the cow, the donkey hit her outside.'					
	l-ʔæb the-father	ed-doktɔ:r the-doctor	fæħasˤ-ɔ examined-him	əs-sɒbəħ the-morning		
	'As for the father, the doctor examined him in the morning.'					
Question (*k* = 6)	mi:n who bə-l-mæktæbe in-the-library	əlli that mbæ:reħ yesterday?	ʃæf-ət-ɔ saw-3SG.F-him	l-bənt the-girl		
	'Who did the girl see [him] in the library yesterday?'					
	mi:n who	ʃæ:f-u have.seen-3PL	jænb near	l-bæ:b the-door	l-ʔæmæ:mi: the-front	
	'Who have they seen near the front door?'					
	ʃu: what	ʃtæra bought	l-ʔæb the-father	eʃ-ʃaher the-month	l-mədˤi the-last	
	'What did the father buy last month?'					
	ʃu: what	tˤabx-et cooked-3SG.F	l-ʔm the-mother	l-mæsæ the-evening		
	'What did the mother cook in the evening?'					
	ʔyyæ which	ʃara:b drink	kæb spilled	l-ja:r the-neighbor	bə-l-be:t in-the-house	
	'Which drink did the neighbor spill in the house?'					
	ʔyyæ which	sˤu:ra picture	rasæm-æ drew-her	bə-l-be:t in-the-house	mbæreħ yesterday	
	'Which picture did he draw [it] at home yesterday?'					

Table A2. *Cont.*

Structure	Example				
Coordinated (*k* = 3)	*l-kælb* the-dog	*ʕæm* PROG	*y-ʕaww-i* 3SG.M-bark-SG	*barra* outside	*w* and
	l-wælæd the-child	*ʕæm* PROG	*y-ibk-i* 3SG.M-cry-SG	*juwwa* inside	
	'The dog is barking outside and the child is crying inside.'				
	l-ʔm the-mother	*ʕæm* PROG	*tət-sæwwaʔ* 3SG.F-shop	*w* and	
	l-wælæd the-boy	*ʕæm* PROG	*yi-drɔs* 3SG.M-study	*bə-l-be:t* in-the-home	
	'The mother is shopping and the boy is studying at home.'				
	ja:r-næ neighbor-our	*ʕæm* PROG	*y-inæd*ˤ*ef* 3SG.M-clean	*es-sæyyæra* the-car	*w* and
	əbn-ɔ son-his	*ʕæm* PROG	*yə-lʕab* 3SG.M-play	*bə-t*ˤ*-t*ˤ *a:be* with-the-ball	
	'Our neighbor is cleaning the car and his son is playing with the ball.'				
Subordinate (*k* = 4)	*fet*ˤ*er* had.breakfast	*l-wælæd* the-child	*bæʕd-mæ: ɣassæl* after washed	*wəʃʃ- ɔ* face-his	
	'The child had breakfast after he washed his face.'				
	ʔabəl-mæ: before	*tə-t*ˤ*aʃʃæ* 3SG.F-has.dinner	*l-bənt* the-girl	*rɒħ* FUT	
	tə-lʕab 3SG.F-play	*ʕæ-l-kɔmbiuter* on-the-computer			
	'Before the girl has dinner, she will play with the computer.'				
	rɒħ FUT	*y-a:xd-u* 3-get-PL	*l-wlæ:d* the-children	*hdi:ə* gift	
	izæ if	*nad*ˤ*d*ˤ*af-u* clean-3PL	*l-be:t* the-house		
	'The children will get a present if they clean the house.'				
	izæ if	*kæ:n* was	*ej-ʒau* the-weather	*dæ:fi* warm	
	fi:-næ can-1PL	*nə-t*ˤ*laʕ* 1PL-go.out	*ʕæ-l-jnene* to-the-park		
	'If the weather is warm, we can go to the park.'				
Relative (*k* = 6)	*ənbasat*ˤ*-u* enjoyed-3PL	*l-wlæ:d* the-children	*bə-ʃ-ʃɔkɔlata* with-the-chocolate		
	əlli that	*ʔækæl-u-wa* ate-3PL-her			
	'The children enjoyed the chocolate that they ate [it].'				
	l-ʔm the-mother	*sæw-et* made-3SG.F	*et*ˤ*-t*ˤ*abxa* the-meal	*yælli* that	
	ʕæm PROG	*y-æ:kl-u-wa* 3-eat-PL-her	*l-wlæ:d* the-children		
	'The mother made the meal that the children are eating [it].'				
	læ:zem should	*y-ħæməm-u* 3-wash-PL	*ez-*ˤ*ɣi:r* the-baby	*əlli* that	
	ʔm-ɔ mother-his	*ʕæm* PROG	*ət-t*ˤ*aʕm-i:* 3SG.F-feed-him		
	'They should wash the baby that his mother is feeding [him].'				
	*l-ħs*ˤ*a:n* the-horse	*əlli* that	*l-fællæħ* the-farmer	*d*ˤ*arab* hit	
	rafæs-ɔ kicked-him	*ʕæ-d*ˤ*ahr-ɔ* on-back-his			
	'The horse that the farmer hit kicked him on his back.'				
	*es*ˤ*-s*ˤ*abi* the-boy	*əlli* that	*sæ:ʕæd-ɔ* help-him	*l-ja:r* the-neighbor	
	*d*ˤ*ayyæʕ* lost	*t*ˤ*ari:ʔ-ɔ* way-his			
	'The boy that the neighbor helped [him] has lost his way.'				
	fæ:z won	*l-fari:ʔ* the-team	*əlli* that	*ʃæjæʕ-ɔ* encouraged-him	
	æx-i brother-my	*bə-s-siba:ʔ* in-the-race			
	'The team that my brother cheered for [it] won the race.'				

Appendix B. Comparisons in Performance at T1 and T2 in English and Arabic

Table A3. Mean (*SD*) morphosyntactic accuracy score by the total sample and by the Younger and Older subgroups separately.

		Total Sample	Older	Younger
English	T1	18.38 (8.32)	19.38 (8.43)	17.34 (8.16)
	T2	23.69 (6.36)	23.94 (6.14)	23.42 (6.63)
Arabic	T1	26.34 (4.53)	27.42 (3.14)	25.22 (5.44)
	T2	27.57 (3.88)	28.83 (1.96)	26.26 (4.86)

Note. Older and Younger are subgroups of the Total sample.

Table A4. Coefficients of the fixed effects of the logistic mixed-effects regression comparing performance in English and Arabic at T1 and T2 for the whole sample.

	Estimate	Std. Error	z Value	Pr (>|z|)
(Intercept)	3.030	0.322	9.405	<0.001 ***
LanguageEnglish	−2.161	0.435	−4.963	<0.001 ***
TimeT2	0.545	0.119	4.585	<0.001 ***
LanguageEnglish:TimeT2	0.673	0.130	5.170	<0.001 ***

Note. The intercept is Arabic at T1, since the reference level for Language is Arabic and for Time it is T1. *** = $p < 0.001$.

Table A5. Random effects of the logistic mixed-effects regression comparing performance in English and Arabic at T1 and T2 for the whole sample.

Groups	Name	Variance	Std. Dev.	Corr	
Participant	(Intercept)	1.722	1.312		
	TimeT2	0.307	0.554	0.00	
	LanguageEnglish	2.192	1.481	−0.23	−0.26
Item	(Intercept)	2.426	1.558		

Table A6. Coefficients of the fixed effects of the logistic mixed-effects regression comparing performance in English and Arabic at T1 and T2 for the two AOA subgroups.

	Estimate	Std. Error	z Value	Pr (>|z|)
(Intercept)	3.312	0.351	9.435	<0.001 ***
AOAYounger	−0.599	0.283	−2.114	0.035 *
TimeT2	0.889	0.178	4.987	<0.001 ***
LanguageEnglish	−2.174	0.468	−4.643	<0.001 ***
AOAYounger:TimeT2	−0.594	0.232	−2.557	0.011 *
AOAYounger:LanguageEnglish	0.056	0.339	0.165	0.869
TimeT2: LanguageEnglish	0.170	0.192	0.886	0.376
AOAYounger:TimeT2:LanguageEnglish	0.907	0.254	3.576	<0.001 ***

Note. The intercept is Arabic at T1 for the older subgroup, since the reference level for Language is Arabic, for Time it is T1, and for AOA it is Older. *** = $p < 0.001$; * = $p < 0.05$.

Table A7. Random effects of the logistic mixed-effects regression comparing performance in English and Arabic at T1 and T2 for the two AOA subgroups.

Groups	Name	Variance	Std. Dev.	Corr	
Participant	(Intercept)	1.550	1.245		
	TimeT2	0.317	0.563	−0.01	
	LanguageEnglish	2.138	1.462	−0.19	−0.29
Item	(Intercept)	2.435	1.560		

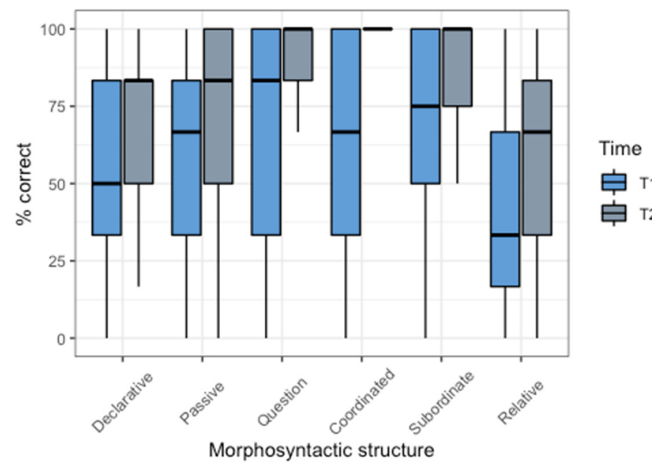## Appendix C. Performance on Individual Structures

Figure A1. Boxplot of participants' performance on each structure of the English SRT at T1 and T2. Percentages of correct responses are shown on the *y*-axis.
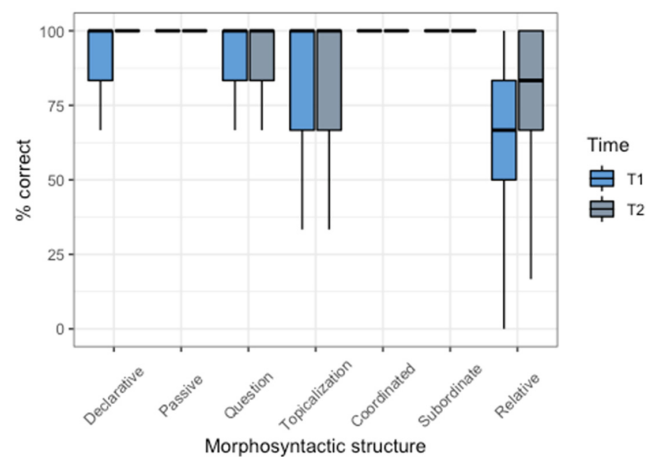
Figure A2. Boxplot of participants' performance on each structure of the Arabic SRT at T1 and T2. Percentages of correct responses are shown on the *y*-axis.

Appendix D. Full Model Outputs and Post-Hoc Tests

Appendix D.1. English T1

Table A8. English T1 fixed effects.

	Estimate	Std. Error	*z* Value	Pr (>|z|)
(Intercept)	1.131	0.527	2.148	0.032 *
English AOA	0.300	0.150	2.005	0.045 *
Length of English schooling	0.361	0.136	2.657	0.008 **
Non-verbal analytical skills	0.368	0.149	2.466	0.014 *
Verbal memory	0.376	0.140	2.689	0.007 **
English environment richness	0.274	0.139	1.969	0.049 *
Language use with siblings	0.534	0.151	3.540	<0.001 ***
Maternal education	0.558	0.138	4.055	<0.001 ***
Structure_Coordinated	0.440	0.879	0.501	0.616
Structure_Relative	−1.451	0.720	−2.015	0.044 *
Structure_Subordinated	0.740	0.805	0.920	0.358
Structure_Passive	−0.386	0.719	−0.538	0.591
Structure_Question	0.181	0.720	0.252	0.801

Note. Reference level for the Structure factor was set to Declarative. *** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.05$.

Table A9. English T1 random effects.

Groups	Name	Variance	Std. Dev.
Participant	(Intercept)	1.343	1.159
Item	(Intercept)	1.450	1.204

Appendix D.2. Syrian Arabic T1

Table A10. Syrian Arabic T1 fixed effects.

	Estimate	Std. Error	z Value	Pr (>|z|)
(Intercept)	3.752	0.470	7.981	<0.001 ***
English AOA	0.325	0.133	2.452	0.014 *
Non-verbal analytical skills	0.124	0.134	0.926	0.354
Verbal memory	0.522	0.127	4.099	<0.001 ***
Structure_Coordinated	0.279	0.796	0.350	0.726
Structure_Relative	−2.775	0.616	−4.503	<0.001 ***
Structure_Subordinated	−0.779	0.695	−1.121	0.262
Structure_Topicalization	−2.002	0.744	−2.691	0.007 **
Structure_Passive	0.587	0.808	0.727	0.467
Structure_Question	−0.362	0.638	−0.566	0.571

Note. Reference level for the Structure factor was set to Declarative. *** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.05$.

Table A11. Syrian Arabic T1 random effects.

Groups	Name	Variance	Std. Dev.
Participant	(Intercept)	0.968	0.984
Item	(Intercept)	0.957	0.978

Table A12. Post-hoc tests with a Tukey adjustment showing pairwise contrasts by morphosyntactic structure in Syrian Arabic T1 model.

Contrast	Estimate	SE	z.Ratio	p.Value
Structure_Declarative—Structure_Coordinated	−0.279	0.796	−0.350	1.000
Structure_Declarative—Structure_Relative	2.775	0.616	4.503	<0.001 ***
Structure_Declarative—Structure_Subordinated	0.779	0.695	1.121	0.922
Structure_Declarative—Structure_Topicalization	2.002	0.744	2.691	0.100
Structure_Declarative—Structure_Passive	−0.587	0.808	−0.727	0.991
Structure_Declarative—Structure_Question	0.362	0.638	0.566	0.998
Structure_Coordinated—Structure_Relative	3.054	0.777	3.930	0.002 **
Structure_Coordinated—Structure_Subordinated	1.058	0.841	1.258	0.871
Structure_Coordinated—Structure_Topicalization	2.281	0.882	2.587	0.130
Structure_Coordinated—Structure_Passive	−0.308	0.936	−0.329	1.000
Structure_Coordinated—Structure_Question	0.640	0.795	0.805	0.985
Structure_Relative—Structure_Subordinated	−1.996	0.670	−2.977	0.046 *
Structure_Relative—Structure_Topicalization	−0.773	0.720	−1.073	0.936
Structure_Relative—Structure_Passive	−3.362	0.790	−4.257	<0.001 ***
Structure_Relative—Structure_Question	−2.413	0.615	−3.924	0.002 **
Structure_Subordinated—Structure_Topicalization	1.223	0.790	1.549	0.715
Structure_Subordinated—Structure_Passive	−1.366	0.852	−1.603	0.680
Structure_Subordinated—Structure_Question	−0.417	0.694	−0.602	0.997
Structure_Topicalization—Structure_Passive	−2.589	0.893	−2.900	0.057 .
Structure_Topicalization—Structure_Question	−1.641	0.743	−2.207	0.291
Structure_Passive—Structure_Question	0.948	0.807	1.175	0.904

Note. *** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.05$, . = <0.10.

Appendix A.3. English T2

Table A13. English T2 fixed effects.

	Estimate	Std. Error	z Value	Pr (>|z|)
(Intercept)	2.031	0.527	3.854	<0.001 ***
Non-verbal analytical skills	0.553	0.157	3.532	<0.001 ***
Verbal memory	0.758	0.153	4.968	<0.001 ***
Language use with siblings	0.392	0.153	2.559	0.010 *
Structure_Coordinated	1.642	0.885	1.856	0.063 .
Structure_Relative	−1.170	0.709	−1.651	0.099 .
Structure_Subordinated	1.176	0.801	1.469	0.142
Structure_Passive	−0.666	0.709	−0.939	0.347
Structure_Question	0.746	0.714	1.045	0.296

Note. Reference level for the Structure factor was set to Declarative. *** = $p < 0.001$; * = $p < 0.05$, . = <0.10.

Table A14. English T2 random effects.

Groups	Name	Variance	Std. Dev.
Participant	(Intercept)	1.716	1.310
Item	(Intercept)	1.387	1.178

Table A15. Post-hoc tests with a Tukey adjustment showing pairwise contrasts by morphosyntactic structure in English T2 model.

Contrast	Estimate	SE	z.Ratio	p.Value
Structure_Declarative—Structure_Coordinated	−1.642	0.885	−1.856	0.430
Structure_Declarative—Structure_Relative	1.170	0.709	1.651	0.564
Structure_Declarative—Structure_Subordinated	−1.176	0.801	−1.469	0.684
Structure_Declarative—Structure_Passive	0.666	0.709	0.939	0.936
Structure_Declarative—Structure_Question	−0.746	0.714	−1.045	0.903
Structure_Coordinated—Structure_Relative	2.813	0.878	3.205	0.017 *
Structure_Coordinated—Structure_Subordinated	0.466	0.951	0.490	0.997
Structure_Coordinated—Structure_Passive	2.308	0.877	2.631	0.090 .
Structure_Coordinated—Structure_Question	0.896	0.880	1.018	0.912
Structure_Relative—Structure_Subordinated	−2.346	0.792	−2.962	0.036 *
Structure_Relative—Structure_Passive	−0.505	0.697	−0.723	0.979
Structure_Relative—Structure_Question	−1.916	0.704	−2.723	0.071 .
Structure_Subordinated—Structure_Passive	1.842	0.792	2.326	0.184
Structure_Subordinated—Structure_Question	0.430	0.795	0.540	0.995
Structure_Passive—Structure_Question	−1.412	0.704	−2.007	0.338

Note. * = $p < 0.05$, . = <0.10.

Appendix A.4. Syrian Arabic T2

Table A16. Syrian Arabic T2 fixed effects.

	Estimate	Std. Error	z Value	Pr (>|z|)
(Intercept)	4.355	0.578	7.529	<0.001 ***
English AOA	0.619	0.143	4.317	<0.001 ***
Non-verbal analytical skills	0.249	0.145	1.718	0.086 .
Verbal memory	0.387	0.126	3.075	0.002 **
Language use with siblings	−0.401	0.125	−3.203	0.001 **
Structure_Coordinated	0.372	0.967	0.385	0.700
Structure_Relative	−2.636	0.747	−3.527	<0.001 ***
Structure_Subordinated	−0.059	0.863	−0.068	0.946
Structure_Topicalization	−2.249	0.896	−2.510	0.012 *
Structure_Passive	0.663	0.990	0.669	0.503
Structure_Question	−0.412	0.775	−0.532	0.595

Note. *** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.05$, . = <0.10.

Table A17. Syrian Arabic T2 random effects.

Groups	Name	Variance	Std. Dev.
Participant	(Intercept)	0.8064	0.898
Item	(Intercept)	1.3606	1.166

Table A18. Post-hoc tests with a Tukey adjustment showing pairwise contrasts by morphosyntactic structure in Syrian Arabic T2 model.

Contrast	Estimate	*SE*	*z*.Ratio	*p*.Value
Structure_Declarative—Structure_Coordinated	−0.372	0.967	−0.385	1.000
Structure_Declarative—Structure_Relative	2.636	0.747	3.527	0.008 **
Structure_Declarative—Structure_Subordinated	0.059	0.863	0.068	1.000
Structure_Declarative—Structure_Topicalization	2.249	0.896	2.510	0.155
Structure_Declarative—Structure_Passive	−0.663	0.990	−0.669	0.994
Structure_Declarative—Structure_Question	0.412	0.775	0.532	0.998
Structure_Coordinated—Structure_Relative	3.008	0.935	3.219	0.022 *
Structure_Coordinated—Structure_Subordinated	0.431	1.029	0.419	1.000
Structure_Coordinated—Structure_Topicalization	2.621	1.057	2.479	0.167
Structure_Coordinated—Structure_Passive	−0.290	1.140	−0.255	1.000
Structure_Coordinated—Structure_Question	0.784	0.960	0.817	0.983
Structure_Relative—Structure_Subordinated	−2.578	0.825	−3.125	0.029 *
Structure_Relative—Structure_Topicalization	−0.387	0.855	−0.453	0.999
Structure_Relative—Structure_Passive	−3.299	0.962	−3.428	0.011 *
Structure_Relative—Structure_Question	−2.224	0.738	−3.013	0.041 *
Structure_Subordinated—Structure_Topicalization	2.191	0.962	2.278	0.255
Structure_Subordinated—Structure_Passive	−0.721	1.054	−0.684	0.993
Structure_Subordinated—Structure_Question	0.353	0.856	0.413	1.000
Structure_Topicalization—Structure_Passive	−2.912	1.082	−2.692	0.100
Structure_Topicalization—Structure_Question	−1.837	0.889	−2.068	0.372
Structure_Passive—Structure_Question	1.074	0.983	1.093	0.930

Note. ** = $p < 0.01$; * = $p < 0.05$.

References

Albirini, Abdulkafi. 2014. Toward understanding the variability in the language proficiencies of Arabic heritage speakers. *International Journal of Bilingualism* 18: 730–65. [CrossRef]

Albirini, Abdulkafi. 2018. The Role of Age of Exposure to English in the Development of Arabic as a Heritage Language in the United States. *Language Acquisition* 25: 178–96. [CrossRef]

Altman, Carmit, Zhanna Feldman Burstein, Dafna Yitzhaki, Sharon Armon-Lotem, and Joel Walters. 2014. Family language policies, reported language use and proficiency in Russian–Hebrew bilingual children in Israel. *Journal of Multilingual and Multicultural Development* 35: 216–34. [CrossRef]

Andreou, Maria, Ifigenia Dosi, Despina Papadopoulou, and Ianthi Maria Tsimpli. 2020. Heritage and non-heritage bilinguals. The role of biliteracy and bilingual education. In *Lost in Transmission. The Role of Attrition and Input in Heritage Language Development*. Edited by Bernhard Brehmer and Jeanine Treffers-Daller. Amsterdam: John Benjamins Publishing Company, pp. 172–96.

Armon-Lotem, Sharon, and Natalia Meir. 2019. The nature of exposure and input in early bilingualism. In *The Cambridge Handbook of Bilingualism*. Edited by Annick De Houwer and Lourdes Ortega. Cambridge: Cambridge University Press, pp. 193–211.

Armon-Lotem, Sharon, Joel Walters, and Natalia Gagarina. 2011. The impact of internal and external factors on linguistic performance in the home language and in L2 among Russian-Hebrew and Russian-German preschool children. *Linguistic Approaches to Bilingualism* 1: 291–317. [CrossRef]

Baayen, R. Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. LME4: Linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]

Bayram, Fatih, Jason Rothman, Michael Iverson, Tanja Kupisch, David Miller, Eloi Puig-Mayenco, and Marit Westergaard. 2017. Differences in use without deficiencies in competence: Passives in the Turkish and German of Turkish heritage speakers in Germany. *International Journal of Bilingual Education and Bilingualism* 22: 919–39. [CrossRef]

Birdsong, David, and Michelle Molis. 2001. On the evidence for maturational constraints in second language acquisition. *Journal of Memory and Language 44*, 235–49. [CrossRef]

Boerma, Tessel, Shula Chiat, Paul Leseman, Mona Timmermeister, Frank Wijnen, and Elma Blom. 2015. A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *Journal of Speech, Language, and Hearing Research* 58: 1747–60. [CrossRef] [PubMed]

Carreira, Maria, and Olga Kagan. 2011. The results of the National Heritage Language Survey: Implications for teaching, curriculum design, and professional development. *Foreign Language Annals* 44: 40–64. [CrossRef]

Chiat, Shula. 2015. Nonword repetition. In *Methods for Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. Edited by Sharon Armon-Lotem, Jan de Jong and Natalia Meir. Bristol: Multilingual Matters, pp. 125–50.

Chiat, Shula, Sharon Armon-Lotem, Theodoros Marinis, Kamila Polišenská, Penny Roy, and Belinda Seeff-Gabriel. 2013. Assessment of language abilities in sequential bilingual children: The potential of sentence imitation tasks. In *Issues in the Assessment of Bilinguals*. Edited by Virginia C. Mueller Gathercole. Bristol: Multilingual Matters, pp. 56–89.

Chondrogianni, Vasiliki, and Theodoros Marinis. 2011. Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguistic Approaches to Bilingualism* 1: 318–45. [CrossRef]

Collins, Brian A., Erin E. O'Connor, Carola Suárez-Orozco, Alfonso Nieto-Castañon, and Claudio O. Toppelberg. 2014. Dual language profiles of Latino children of immigrants: Stability and change over the early school years. *Applied Psycholinguistics* 35: 581–620. [CrossRef] [PubMed]

Cowell, Mark W. 1964. *A Reference Grammar of Syrian Arabic*. Washington: Georgetown University Press.

Daskalaki, Evangelia, Elma Blom, Vasiliki Chondrogianni, and Johanne Paradis. 2020. Effects of Parental Input Quality in Child Heritage Language Acquisition. *Journal of Child Language* 47: 709–36. [CrossRef] [PubMed]

Daskalaki, Evangelia, Vasiliki Chondrogianni, Elma Blom, Froso Argryri, and Johanne Paradis. 2019. Input effects across domains: The case of Greek subjects in child heritage language. *Second Language Research* 35: 421–55. [CrossRef]

DeKeyser, Robert. 2012. Interactions between individual differences, treatments, and structures in SLA. *Language Learning* 62: 189–200. [CrossRef]

Flores, Cristina, Ana Lucía Santos, Alice Jesus, and Rui Marques. 2017. Age and input effects in the acquisition of mood in Heritage Portuguese. *Journal of Child Language* 44: 795–828. [CrossRef] [PubMed]

Golberg, Heather, Johanne Paradis, and Martha Crago. 2008. Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics* 29: 41–65. [CrossRef]

Govindarajan, Krithika, and Johanne Paradis. 2019. Narrative abilities of bilingual children with and without developmental language disorder (SLI): Differentiation and the role of age and input factors. *Journal of Communication Disorders* 77: 1–16. [CrossRef] [PubMed]

Graham, Hamish R., Ripudaman S. Minhas, and Georgia Paxton. 2016. Learning problems in children of refugee background: A systematic review. *Pediatrics* 137: e20153994. [CrossRef] [PubMed]

Hamann, Cornelia, Solveig Chilla, Lina Abed Ibrahim, and István Fekete. 2020. Language assessment tools for Arabic-speaking heritage and refugee children in Germany. *Applied Psycholinguistics* 41: 1375–414. [CrossRef]

Hammer, Carol S., Eugene Komaroff, Barbara L. Rodriguez, Lisa M. Lopez, Shelley E. Scarpino, and Brian Goldstein. 2012. Predicting Spanish–English bilingual children's language abilities. *Journal of Speech, Language, and Hearing Research* 55: 1251–64. [CrossRef]

Hammer, Carol S., Shelley Scarpino, and Megan Dunn Davison. 2011. Beginning with language: Spanish–English bilingual preschoolers' early literacy development. In *Handbook of Early Literacy Research*. Edited by Susan B. Neumann and David K. Dickinson. New York: Guilford Press, vol. 3, pp. 118–35.

Harrell, Frank E., Jr. 2020. Hmisc: Harrell Miscellaneous. R Package Version 4.4-1. Available online: https://www.rdocumentation.org/packages/Hmisc/versions/4.4-1 (accessed on 15 March 2021).

Hoff, Erika, and Krystal M. Ribot. 2017. Language growth in English monolingual and Spanish-English bilingual children from 2.5 to 5 years. *The Journal of Pediatrics* 190: 241–45. [CrossRef] [PubMed]

Hoff, Erika, Cynthia Core, Silvia Place, Rosario Rumiche, Melissa Señor, and Marisol Parra. 2012. Dual language exposure and early bilingual development. *Journal of Child Language* 39: 1–27. [CrossRef] [PubMed]

Hosmer, David W., Jr., Stanley Lemeshow, and Rodney X. Sturdivant. 2013. *Applied Logistic Regression*. Hoboken: John Wiley and Sons, vol. 398.

IRCC. 2020. Resettled Refugees. Available online: http://www.cic.gc.ca/opendata-donneesouvertes/data/IRCC_Resettled_0012_E.xls (accessed on 1 November 2020).

Jia, Gisela. 2008. Heritage language development, maintenance and attrition among recent Chinese immigrants in New York City. In *Chinese as a Heritage Language: Fostering World Citizenry*. Edited by A.W. He and Y. Xiao. Honolulu: University of Hawai'i, National Foreign Language Resource Center, pp. 189–203.

Jia, Gisela, and Doris Aaronson. 2003. A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics* 24: 131–61. [CrossRef]

Jia, Gisela, and Akiko Fuse. 2007. Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: Age-related differences. *Journal of Speech, Language, and Hearing Research* 50: 1280–99. [CrossRef]

Jia, Gisela, Jennifer Chen, HyeYoung Kim, Phoenix-Shan Chan, and Changmo Jeung. 2014. Bilingual lexical skills of school-age children with Chinese and Korean heritage languages in the United States. *International Journal of Behavioral Development* 38: 350–58. [CrossRef]

Jia, Ruiting, and Johanne Paradis. 2015. The use of referring expressions in narratives by Mandarin heritage language children and the role of language environment factors in predicting individual differences. *Bilingualism* 18: 737–52. [CrossRef]

Jia, Ruiting, and Johanne Paradis. 2020. The acquisition of relative clauses by Mandarin heritage language children. *Linguistic Approaches to Bilingualism* 10: 153–83. [CrossRef]

Johnson, Jaqueline S., and Elissa L. Newport. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21: 60–99. [CrossRef]

Kaltsa, Maria, Alexandra Prentza, and Ianthi Maria Tsimpli. 2019. Input and literacy effects in simultaneous and sequential bilinguals: The performance of Albanian–Greek-speaking children in sentence repetition. *International Journal of Bilingualism* 24: 159–183. [CrossRef]

Kaplan, Ida, Yvonne Stolk, Madeleine Valibhoy, Alan Tucker, and Judy Baker. 2016. Cognitive assessment of refugee children: Effects of trauma and new language acquisition. *Transcultural Psychiatry* 53: 81–109. [CrossRef] [PubMed]

Klem, Marianne, Monica Melby-Lervåg, Bente Hagtvet, Solveig-Alma Halaas Lyster, Jan-Eric Gustafsson, and Charles Hulme. 2015. Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science* 18: 146–54. [CrossRef] [PubMed]

Kohnert, Kathryn J., and Elizabeth Bates. 2002. Balancing bilinguals II: Lexical comprehension and cognitive processing in children learning Spanish and English. *Journal of Speech, Language and Hearing Research* 45: 347–59. [CrossRef]

Kohnert, Kathryn. J., Elizabeth Bates, and Arturo E. Hernandez. 1999. Balancing bilinguals: Lexical-semantic production and cognitive processing in children learning Spanish and English. *Journal of Speech, Language, and Hearing Research* 42: 1400–13. [CrossRef] [PubMed]

Komeili, Mariam, Theodoros Marinis, Parvaneh Kavakoli, and Yalda Kazemi. 2020. Sentence Repetition in Farsi-English Bilingual Children. *Journal of the European Second Language Association* 4: 1–12. [CrossRef]

Lenth, Russell V. 2021. Emmeans: Estimated Marginal Means, Aka Least-Squares Means. R Package Version 1.5.4. Available online: https://cran.r-project.org/web/packages/emmeans/index.html (accessed on 15 March 2021).

Lüdecke, Daniel. 2021. _sjPlot: Data Visualization for Statistics in Social Science_. R Package Version 2.8.7. Available online: https://cran.r-project.org/web/packages/sjPlot/citation.html (accessed on 15 March 2021).

Marinis, Theo, and Sharon Armon-Lotem. 2015. Sentence repetition. In *Methods for Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. Edited by Sharon Armon-Lotem, Jan de Jong and Natalia Meir. Bristol: Multilingual Matters.

Meir, Natalia, Joel Walters, and Sharon Armon-Lotem. 2017. Bi-directional cross-linguistic influence in bilingual Russian-Hebrew children. *Linguistic Approaches to Bilingualism* 7: 514–53. [CrossRef]

Montrul, Silvina. 2016. *The Acquisition of Heritage Languages*. Cambridge: Cambridge University Press.

Naglieri, Jack A. 1985. *Matrix Analogies Test: Short Form (MAT-SF)*. San Antonio: Psychological Corporation.

Paradis, Johanne. 2011. Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism* 1: 213–37. [CrossRef]

Paradis, Johanne, Adriana Soto-Corominas, Xi Chen, and Alexandra Gottardo. 2020a. How language environment, age and cognitive capacity support the bilingual development of Syrian refugee children recently arrived in Canada. *Applied Psycholinguistics*. [CrossRef]

Paradis, Johanne, Adriana Soto-Corominas, Xi Chen, and Alexandra Gottardo. 2020b. The home language environment and English second language learning of Syrian refugee children: Comparisons with other newcomer children in Canadian schools. In *Refugees in Canada and Germany: From Research to Policies and Practice*. Edited by Annette Korntheuer, Paul Pritchard, Débora B. Maehler and Lori Wilkinson. Köln: GESIS—Leibniz-Institut für Sozialwissenschaften. [CrossRef]

Paradis, Johanne, Brian Rusk, Tamara Sorenson Duncan, and Krithika Govindarajan. 2017. Children's Second Language Acquisition of English Complex Syntax: The Role of Age, Input, and Cognitive Factors. *Annual Review of Applied Linguistics* 37: 148–67. [CrossRef]

Paradis, Johanne, Fred Genesee, and Martha Crago. 2021. *Dual Language Development and Disorders: A Handbook on Bilingualism and Second Language Learning*, 3rd ed. Baltimore: Brookes Publishing Company.

Pham, Giang, and Kathryn Kohnert. 2014. A longitudinal study of lexical development in children learning Vietnamese and English. *Child Development* 85: 767–82. [CrossRef] [PubMed]

Place, Silvia, and Erika Hoff. 2016. Effects and noneffects of input in bilingual environments on dual language skills in 2$\frac{1}{2}$-year-olds. *Bilingualism* 19: 1023–41. [CrossRef]

Pham, Giang, and Timothy Tipton. 2018. Internal and external factors that support children's minority first language and English. *Language, Speech, and Hearing Services in Schools* 49: 595–606. [CrossRef] [PubMed]

Polišenská, Kamila, Shula Chiat, and Penny Roy. 2015. Sentence repetition: What does the task measure? *International Journal of Language and Communication Disorders* 50: 106–18. [CrossRef]

Prevoo, Marielle J. L., Maike Malda, Judi Mesman, Rosanneke A. Emmen, Nihal Yeniad, Marinus H. van Ijzendoorn, and Marielle Linting. 2014. Predicting ethnic minority children's vocabulary from socioeconomic status, maternal language and home reading input: Different pathways for host and ethnic language. *Journal of Child Language* 41: 963–84. [CrossRef] [PubMed]

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, Available online: https://www.R-project.org/ (accessed on 15 March 2021).

Rojas, Raúl, Aquiles Iglesias, Ferenc Bunta, Brian Goldstein, Claude Goldenberg, and Leslie Reese. 2016. Interlocutor differential effects on the expressive language skills of Spanish-speaking English learners. *International Journal of Speech-Language Pathology* 18: 166–77. [CrossRef] [PubMed]

Silva-Corvalán, Carmen. 2014. *Bilingual Language Acquisition: Spanish and English in the First Six Years*. Cambridge: Cambridge University Press.

Sirin, Selcuk R., and Lauren Rogers-Sirin. 2015. *The Educational and Mental Health Needs of Syrian Refugee Children*. Washington: Migration Policy Institute.

Sorenson Duncan, Tamara, and Johanne Paradis. 2020a. How does maternal education influence the linguistic environment supporting bilingual language development in child second language learners of English? *International Journal of Bilingualism* 24: 46–61. [CrossRef]

Sorenson Duncan, Tamara, and Johanne Paradis. 2020b. Home language environment and children's second language acquisition: The special status of input from older siblings. *Journal of Child Language* 47: 982–1005. [CrossRef] [PubMed]

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Winter, Bodo. 2019. *Statistics for Linguists: An Introduction Using R*. New York: Routledge.

Yip, Virginia, and Stephen Matthews. 2007. *The Bilingual Child*. Cambridge: Cambridge University Press.