CAMBRIDGE
UNIVERSITY PRESS

CrossMark

RESEARCH ARTICLE

Lexical diversity development in newly arrived parent-child immigrant pairs: Aptitude, age, exposure, and anxiety

Amelia Lambelet[1,2]

[1]University of Fribourg, Switzerland and [2]CUNY Hunter College, USA
Email: amelia.lambelet@unifr.ch

Abstract

The Language Aptitude Outside the Classroom (LAOC) study investigates the factors that contribute to successful English-learning among newly arrived parent-child immigrants. Two types of factors are considered: cognitive abilities (aptitude measured with the LLAMA tests and working memory) and contextual-affective factors (exposure and anxiety). Participants are pairs of Spanish-speaking immigrants in the US. Each pair consists of a parent and their child aged 7–16. Their English proficiency is measured longitudinally during a one-year period using a listening comprehension test, a verbal fluency test, and an oral narrative (frog story). This contribution focuses on the lexical diversity of the oral narratives (Guiraud Index). Linear mixed models were run on the entire sample and on adults and children separately using time, aptitude, working memory, exposure to English, and anxiety as predictors of lexical diversity of the oral narratives (random effect = dyad, random slope = time). The results show that the development of lexical diversity over a one-year period is predicted by exposure to the language (and, for the children, anxiety). Two subtests of the LLAMA aptitude battery are also significant predictors when the entire sample is considered, but this effect nevertheless disappears for the adults when modeled separately from the children.

Keywords: Aptitude; Working memory; Exposure; Anxiety; Age factor

Theoretical and Empirical Background

According to the latest statistics published by the United Nations, in 2019, 3.5% of the global population, i.e., 272 million people, were international migrants (UN DESA, 2019). Many of these migrants have settled in a country where their first language (L1) is not the majority language and are therefore in the process of acquiring a second language in immersion. This process and its results in terms of the development of language proficiency are mediated by a variety of factors, including individual and contextual differences. In the following, we briefly review four factors: language learning aptitude, age of onset, exposure to the majority language, and anxiety.

Foreign Language Learning Aptitude and Second Language Proficiency Development

Foreign language aptitude has been found to be an important predictive factor of success in developing language proficiency since Carroll's pioneer work in the 1960s on the abilities needed to learn foreign languages in short and intense foreign language training courses (Carroll, 1962, 1973). His research resulted in the definition of foreign language aptitude as consisting of four dimensions: *phonetic coding ability* (the ability to discriminate and identify new language sounds), *grammatical sensitivity* (the awareness of the grammatical function of the different elements constituting a sentence), *inductive ability* (the ability to identify grammatical/meaning patterns in an unknown language sample), and *rote memory* (the ability to learn a large number of items in a short time). Since this definition, foreign language aptitude has demonstrated its predictive power in a number of studies aiming to investigate inter-individual differences in foreign language learning in school contexts (among others Doughty et al., 2010; Erlam, 2005; Kiss & Nikolov, 2005; see also Li, 2015 for a meta-analysis).

The importance of Carroll's four dimensions of aptitude has nevertheless been critiqued both theoretically and empirically over the years. Skehan (2002), for instance, proposed a model of aptitude using three dimensions (auditory processing, language analysis, and memory), while Robinson (2005, 2007, 2012) introduced the notion of aptitude complexes, or sets of abilities activated differently depending on the stage of acquisition the learner is in.

Another critique is directed at Carroll's definition of (rote) memory as the "ability to store information passively" (Erlam, 2005, p. 149). This conceptualization of memory as static and passive has been challenged since Baddeley & Hitch (1974) defined working memory as consisting of three components (the central executive, the phonological loop, and the visuospatial sketchpad) that not only store information but also process it in real-time, and serve as a gateway to long term memory. Several recent studies have found working memory to "exercise consistent and distinctive influences on various aspects of L2 acquisition and processing" (Wen et al., 2016, p. 19). Given the range of findings, the importance of working memory deserves more empirical testing (DeKeyser, 2019; Li, 2015; Sáfár & Kormos, 2008; Singleton, 2014; Wen, 2019; Wen et al., 2016; Wen & Skehan, 2011).

Another issue concerning aptitude is its role in naturalistic settings. Krashen (1981), for instance, claimed that aptitude (or at least its analytic language ability component) is not a relevant factor in naturalistic second language acquisition settings (i.e., in an immigration context) where other factors might play a greater role. Other scholars have, on the contrary, claimed that aptitude could be even more relevant in immersive than school contexts, as naturalistic learners must find regularities in a great amount of input (see Granena, 2014; Skehan, 1991). Furthermore, in Li's (2015) meta-analysis, where six out of the 33 studies that were considered investigated aptitude in naturalistic settings, it appeared that "the mean effect size associated with naturalistic learning was also significant, which seemed to suggest that aptitude was drawn on in untutored contexts as well as in language classes." (p. 398)

Age of Onset and Second Language Proficiency Development

Aptitude has also been investigated in relation to another factor commonly considered as predicting language proficiency development: biological age of onset. Following Granena & Long (2013b), for instance, age of onset would typically explain about 30% of the variance in second language ultimate attainment; younger learners usually

achieve a higher level of proficiency than older learners in the long run. Note, however, that this advantage of young starters only shows up after a significant amount of exposure to the language. During the first months of language acquisition (i.e., in terms of rate of acquisition), the tendency goes in the opposite direction: older learners usually develop proficiency in the second language faster than younger learners (Asher & Price, 1967; Krashen et al., 1979; Snow & Hoefnagel-Hohle, 1978).

Over the last decades, several maturational hypotheses have been proposed to explain this (paradoxical) effect of age on second language acquisition (see Lambelet & Berthele, 2015; Singleton, 2005; Singleton & Ryan, 2004 for a discussion). One hypothesis concerns procedural differences between adults and children. According to this view of maturational constraints on language learning, children can acquire languages implicitly, while adults need more explicit teaching and learning to develop proficiency because their implicit learning ability is diminished (DeKeyser, 1994, 2008, 2019; DeKeyser & Larson-Hall, 2005; Ellis, 2005, 2014). Since a greater amount of exposure is needed to acquire regularities without explicit teaching, the use of implicit learning skills could explain the disadvantage of children over adults at the beginning stage of L2 acquisition. On the other hand, adults and adolescents would rarely achieve full proficiency in the target language because of the conscious and laborious efforts that explicit learning requires.

Age of Onset, Aptitude, and Second Language Proficiency Development

Even though children usually outperform adults in the long term, some adults achieve a level of proficiency high enough to be considered native-like. From a maturational perspective, these highly proficient late L2 starters would have, for some reason, kept some of their initial language acquisition ability, and therefore would show a higher level of foreign language learning aptitude (Carroll, 1973; Doughty et al., 2010; Ross et al., 2002). In this view, although children acquiring a foreign or second language early in life would not need any special ability to develop the language, only gifted adolescents and adults would be able to achieve a high L2 level. Research results of several studies on native-like L2 users seem to confirm this hypothesis. For instance, in a study with 56 Hungarians who immigrated to the United States either as adults (late starters) or before the age of 16 (early starters), foreign language aptitude appeared as a predictor of English proficiency in late starters *but not* in early starters (DeKeyser, 2000). Similarly, in a second study with immigrants in the United States and in Israel, correlations were found between foreign language aptitude and English proficiency scores for participants who began their L2 learning between the ages of 18 and 40, but not for early starters or those who began their learning after age 40 (DeKeyser et al., 2010).

More nuanced results appear in Granena & Long (2013a). They investigated the predictive effect of age, length of stay in the L2 country and foreign language aptitude on L2 language proficiency in 65 Chinese immigrants in Spain. While Granena & Long (2013a) did find predicted correlations between aptitude and phonology and between aptitude and lexis in the late starter group but not in the early starters, they did not find correlations between aptitude and morphosyntax in any group. As a whole, these results suggest that in immersive settings, aptitude does not play an important role in young learners. In contrast, however, Abrahamsson & Hyltenstam (2008) found an effect of aptitude in *both* early and late starters in a study on 42 Chilean immigrants in Sweden. One possible explanation for these contradictory results is that the measure of L2 proficiency used by Abrahamsson & Hyltenstam was more discriminating than the measures of L2 proficiency used in the other studies. This could have helped to avoid a ceiling effect in the early starter groups that would have "[made] it harder to show the influence of any

moderator variable, including aptitude, and easier to show its effects in adult groups with the greater variability in ultimate attainment that is typical of late starters" (Granena & Long, 2013a, pp. 333–334). If this explanation is correct, the effect of aptitude in early second language learners should appear in the early phases of language acquisition, when variability in language outcome is still present. Research is therefore needed on the effect of aptitude on rate of acquisition, e.g., with adults and children who have recently arrived in a country where the majority language is different from their L1.

Exposure and Other Factors Correlated with Age

Other studies have emphasized factors correlated with age that could also provide an explanation for the long-term advantage of children over adults, such as length of stay, L2 education (especially if children study the L2 at school and learn to read in it while their parents do not), and L1 use (Moyer, 2011). Studies have also highlighted the importance of affective and exposure variables such as frequency of contact with native speakers as predictors of L2 ultimate attainment (Flege & Liu, 2001; Kinsella & Singleton, 2014; Moyer, 2011).

Based on the results of her study on native-like accent in foreign students in US universities, Moyer (2011) argues that *quality* of exposure (in the sense of exposure where the "learner interacts in functionally significant ways, representing interpersonal as well as instrumental purpose", p. 195) is more important than *quantity* of exposure (length of residence, years of instruction, etc.). In the same vein, the fact that children often encounter more diversified input and come into contact with a larger number of target language speakers in a variety of settings (Flege, 1987) could partially explain their superior learning outcomes in the long term.

Recent research on adult immigrants in various contexts has proposed language anxiety as another important factor that mediates both exposure to the language and development of language proficiency (Dewaele & Sevinç, 2017; Garcia de Blakeley et al., 2017; Sevinç, 2018; Sevinç & Backus, 2017), which confirms the well-documented effect of foreign language anxiety in the classroom (for a review, see Gkonou et al., 2017). Level of foreign language anxiety in turn depends on age of onset, according to a study with multilingual adults (Dewaele et al., 2008), but research is needed on the interaction between age, anxiety, and rate of acquisition in an immersive context.

As in other subfields of SLA, research on age, aptitude, and other individual differences has measured the learning outcomes in many ways. For the present study, the main criteria for the choice of the tests of English proficiency were their suitability for different age groups (see Lambelet & Berthele, 2015, for a discussion of research designs in studies on age effect on L2 learning), and their closeness to real communication skills. In this sense, elicited oral narratives based on picture books or short video sequences allow participants to make use of their linguistic repertoire to resolve the task as they would do in real-life situations, while ensuring the comparability across participants thanks to the common thematic frame of the story to be told. Elicited narratives have been used in a wide range of SLA and heritage language studies with adults and children to measure language dominance, narrative complexity, verbal tense production, motion events expressions, or lexical diversity (e.g., De Clercq & Housen, 2019; Sánchez Abchi et al., 2017; Strömqvist & Verhoeven, 2004; Treffers-Daller, 2009). In this article, we focus on the lexical diversity of oral narratives, a measure that gives an insight on the size of the active vocabulary of a L1 or L2 user and reflects learners' level of language proficiency and the complexity of their developing vocabulary (e.g. Malvern et al., 2004). Lexical diversity has been of interest to researchers since the 1930's,

initially in relation to monolingual first language acquisition and then in relation to bilingual acquisition and second language learning. Over the years, various indices have been developed to measure lexical diversity while avoiding the confounding effect of text length (see Jarvis, 2013, for a discussion). For this contribution, we will use the Guiraud Index of lexical diversity, a measure that has been found to be a good predictor of the human perception of lexical richness (e.g., Vanhove et al., 2019) and is widely used in SLA studies.

The Language Aptitude Outside the Classroom (LAOC) Study

In the present study, we use a longitudinal design to investigate the development of lexical diversity in recently arrived parent-child immigrant pairs. We investigate the effects of age of onset, cognitive abilities (foreign language learning aptitude, working memory), exposure, and anxiety. We aim to answer the following research questions about recently arrived Spanish-speaking immigrants in the United States:

Research question 1: What factors predict rate of English acquisition, measured as the lexical diversity of oral narratives?

Research question 2: Are the same factors predictive of rate of English acquisition in adults as in children?

Methodology

All the participants were tested by the principal investigator of the project who met with each dyad three times over a one-year period. Data collections took place either at participants' homes, in a library, or in a community center. During the first half of the first session (T1), participants completed the LLAMA aptitude tests and two short-term/working memory tests (see below, Tasks). The experimenter gave the instructions orally in Spanish to both adult and child simultaneously and answered any questions the participants had. When she felt it necessary, the experimenter asked the child or the adult to repeat the instructions in their own words to check for misunderstandings. The administration of the six cognitive tests (four aptitude tests and two short-term/working memory tests) took approximately 40 to 50 minutes.

In the second half of the first session, child and adult English proficiency was assessed with three tests (a verbal fluency task, an oral narrative, and a listening comprehension task). While participant 1 (the adult) was answering a listening comprehension task on a laptop computer with headphones, participant 2 (the child) performed the verbal fluency and oral narrative (frog story) tasks with the experimenter. After reversing the tasks (the child using the laptop for the listening comprehension task, the adult with the experimenter for the oral tasks), parent and child answered a short questionnaire in Spanish about their exposure to English, their anxiety when speaking in English and other socio-demographic questions. At T2 and T3, respectively 6 months and 12 months after the first session, participants' English proficiency was assessed a second and third time with a variation of the same tests. Participants also answered the socio-biographic questionnaire. The first session lasted for 90 to 120 minutes. Sessions 2 and 3 lasted for 45 to 60 minutes. The study was approved by the IRB of the University of Maryland.

Participants

Participants are 38 parent-child dyads of Spanish-speaking immigrants in New York city ($N = 76$). An additional 13 parent-child dyads were excluded from the analysis because they did not participate in all three data collections or had missing answers

(total participant sample = 51 dyads). Table 1 shows the gender and mean age by group of the participants included in the analysis. Participants come from 9 Latin American countries reflecting the current waves of immigration in the United States: Venezuela ($N = 28$), Dominican Republic ($N = 22$), Ecuador ($N = 8$), Honduras ($N = 6$), Mexico ($N = 4$), Bolivia ($N = 2$), El Salvador ($N = 2$), Peru ($N = 2$) and Puerto Rico ($N = 2$). The answers of the participants to a brief socio-biographical questionnaire shows that the majority of them arrived in the United States with a very low level of proficiency: 1 adult and 5 children reported not understanding a single word when they entered the country; 27 adults and 22 children reported knowing a few words such as "good morning," "thank you," some numbers and some colors; 10 adults and 9 children reported being able to communicate on very basic everyday needs; 1 child but none of the adults reported being able to participate in a conversation on everyday topics; and 1 child (but no adults) reported being able to communicate with certain fluency on a variety of topics.

Tasks

Foreign Language Learning Aptitude Tests

Participants' foreign language learning aptitude was assessed with the four subtests of the LLAMA aptitude tests (Meara, 2005). The LLAMA tests are computer-run and picture-based exercises simulating the learning of an artificial language. They can be used by speakers from any L1 because they do not rely on any specific language system, and their user-friendly interface makes them easily usable for any type of population (including children) which explains their wide use in the field of aptitude, even if their internal validity has been recently questioned (Bokander & Bylund, 2020). The first test, LLAMA_B simulates vocabulary learning. The participant has 120 seconds to learn the names of a set of invented objects (drawings) in an unknown language (training phase). They are then tested on their learning. LLAMA_D is a sound discrimination and recognition task. It is intended to measure the participant's ability to recognize oral patterns in an unknown language. Participants first hear a series of sounds (training phase) and then must discriminate between new and previously heard items (test phase). LLAMA_E measures sound-symbol association ability. For 120 seconds, participants learn the relationship between 22 sounds and 22 written forms in an artificial orthographic system (training phase). They then hear a word and choose the correct written word between two variants (test phase). LLAMA_F measures inductive learning ability. For 300 seconds, participants infer the grammatical system of an artificial language with a set of visual and written stimuli (training phase). They then choose the grammatically correct variant out of two new stimuli (test phase).

Working Memory and Short-Term Memory Tests

Participants answered two tests of working/short term memory on two different laptops with headphones. The tests were built on the PEBL environment (Mueller & Piper, 2014). Participants began with the Corsi Block task, which measures visuo-spatial short-term memory. Participants observed sequences of blocks light up on their screen, and then repeated the sequences back in order. The sequence increased in length at each successful trial until the participant was unable to remember it. When they made errors, participants were given a second chance with a new sequence of the same length before continuing to the increased lengths. After two missed trials, the task was terminated, and the score saved. The second working memory test was the backward digit span.

Table 1. Participants' gender and age at T1 and at arrival by group.

Group	Age at T1			Age on arrival			Gender	
	Mean (*SD*)	Min.	Max.	Mean (*SD*)	Min.	Max.	Females	Males
Adults (*N* = 38)	37 (7)	25	52	38 (7)	26	58	34	4
Children (*N* = 38)	11 (3)	7	16	9 (3)	4	14	21	17
Total (*N* = 76)							55	21

Note: the age on arrival is estimated based on the participants' age at T1 (in years) and their length of residence at T1 (in months).

Participants saw sequences of single digits and were asked to tap the sequence in the reverse order than they appeared on the screen. As for the Corsi Block task, sequences of digits increased in size at each trial. After two missed trials, the score was saved. Corsi block and backward digit span were chosen to avoid confounding memory and language abilities and because they could be administrated to adults and children simultaneously.

Oral Narrative

At each session, adults and children were asked to perform an oral narrative task using the Frog Stories books by Mercer Mayer. At T1 and T3, participants told the story illustrated in the picture book *Frog, where are you?* (Mayer, 1969), while at T2 they told the story illustrated in the picture book *A boy, a dog, and a frog* (Mayer, 1967). Before telling the story, they were given time to look through the book and prepare mentally. Preparation and performance were untimed, and performance was recorded. Here, we only consider the oral narratives from T1 and T3 since T2 pictured a different story, therefore not allowing for direct comparison.

Socio-Biographic Questionnaire

At the end of each session (T1, T2, and T3), participants individually answered a short questionnaire in Spanish. The experimenter helped in reading the questions when necessary. The adult version of the questionnaire contained nine questions on exposure, six questions on anxiety when speaking in English, five questions on motivation and feeling of integration and three questions on education level, employment in the country of origin and employment in the United States. The exposure, anxiety, motivation and feeling of integration questions were adapted for the child questionnaire.

Exposure to English was self-rated in eight domains: at home, with friends, in school/at work, in the neighborhood, in church, when reading, when watching television, and when listening to music. For each of these domains, the participant answered if they were using *only English, mainly English, as much English as Spanish, mainly Spanish,* or *only Spanish.* At T2 and T3, anxiety was self-rated in six domains: at home, with friends, in the neighborhood, in school/at work, on the phone, during adult ESL classes (last question only in the adult questionnaire). Participants answered if they were nervous in each context on a five-point scale ranging from *not nervous at all* to *extremely nervous.* At T1, the questionnaire contained just one general question on anxiety when speaking in English (same five-point scale).

Data Analysis
Data Preparation
Cognitive Variables

The scores for the LLAMA_B, LLAMA_E, and LLAMA_F were retrieved directly from the software. The possible scores on the three subtests range from 0 to 100. The LLAMA_D score as computed from the software ranges from 0 to 75. Participants' scores on the LLAMA_D were therefore transformed to correspond to the range of the other subtests (0–100). The Backward Digit and Corsi Block Spans were retrieved directly from PEBL. To run the linear mixed models, each of the cognitive variables was mean centered (z_scores of the combined age groups for the first model, then separately by age group for the second part of the analysis (see the following sections below: *Factors Explaining Lexical Diversity Development in the Entire Sample (RQ 1)*

and *Factors Explaining Lexical Diversity Development in Adults and Children Separately* (*RQ 2*)).

Affective and Contextual Variables

The answers to the eight questions of the questionnaire on exposure were coded on a scale from 1 to 5 (1 = *only Spanish*, 5 = *only English*), averaged for the three sessions, and transformed to a score of 1 to 100. The same applied to the questions on anxiety. The internal consistency of the *Exposure* and *Anxiety* variables was very satisfactory (Exposure to English: Cronbach's alpha = .92, $N = 24$; Anxiety when speaking in English: Cronbach's alpha = .86, $N = 13$). To run the linear mixed models, the newly created *Exposure* and *Anxiety* variables were centered (z_scores of the combined age groups for the first model, then separately by age group for the second part of the analysis, (see sections below: *Factors Explaining Lexical Diversity Development in the Entire Sample* (*RQ 1*) and *Factors Explaining Lexical Diversity Development in Adults and Children Separately* (*RQ 2*)).

Lexical Diversity of the Oral Narratives

Participants' oral narratives were transcribed by the principal investigator as soon as possible after data collection. All transcriptions were then checked and revised by a research assistant who is a native English speaker. Differences in transcriptions were reviewed by the principal experimenter and, if needed, discussed with the assistant. Filled pauses, unintelligible words, and Spanish code-switching were removed from each transcription before the Guiraud Index measure of lexical diversity (number of types divided by the square root of number of tokens) was computed.

Results

Dependent Variable: Guiraud Index of Lexical Diversity

At T1, the mean Guiraud Index score of the children was 4.17 (*SD* = 1.5) and the parents' score was 3.74 (*SD* = 1.1). One year later (T3), the Guiraud Index score of the children increased to 4.96 (*SD* = 1.2) and the parents' score slightly increased to 3.91 (*SD* = 1.1). A 2 × 2 repeated measures ANOVA revealed a significant main effect of group ($F(1, 35) = 16.25$, $p < .001$, $\eta_p^2 = .32$), a significant main effect of time ($F(1, 35) = 22.97$, $p < .001$, $\eta_p^2 = .40$) and a significant effect of the interaction between time and group ($F(1,35) = 10.1$, $p < .001$, $\eta_p^2 = .22$). In other words, (a) the children scored higher than their parents in general, (b) the scores of the entire sample increased between T1 and T3, but (c) this increase differed by group: the difference between T1 and T3 is significant in the child group ($t(35) = 5.30$, $p < .001$), but not in the adults ($t(35) = 1.64$, $p = .11$). The scores of the adults and the children are plotted in Figure 1, below.

Independent Variables

The means and standard deviations for both groups on the nine independent variables are shown in Table 2. Regarding cognitive variables, adults and children differ significantly on the LLAMA_B and LLAMA_F (higher score for the children than for their parents) but not on the other four cognitive tests. They nevertheless differ significantly on both exposure and anxiety; children report more exposure to English in their everyday life than their parents do and report less anxiety when speaking in English. The two groups do not differ significantly in terms of length of residence in the United States.
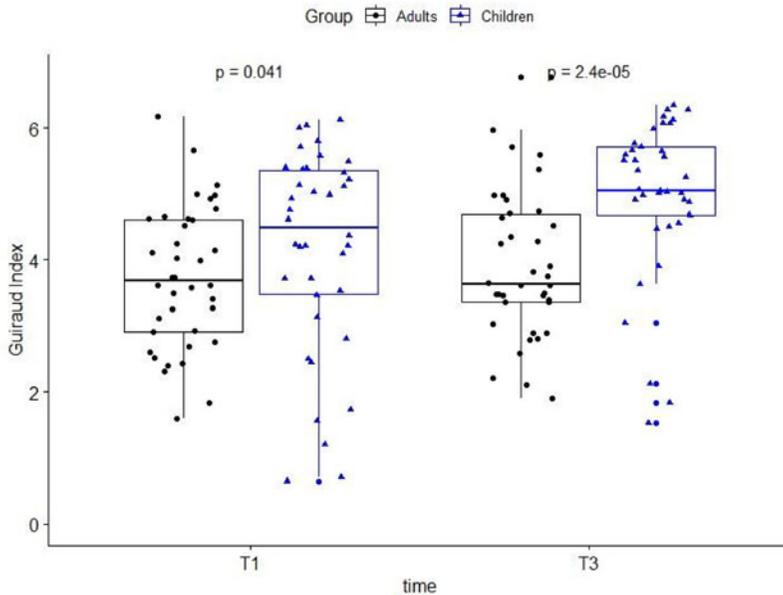
Figure 1. Boxplots: Distribution of the Guiraud Index Scores for Each Group at Each Session (T1 and T3)

Correlation Matrix

Before running linear mixed effects models on the data, we examined the correlations between the Guiraud Index and the cognitive, exposure, and anxiety variables. As shown in Table 3, the Guiraud Index was moderately correlated with the four measures of aptitude, length of residence, exposure and anxiety. Contrary to expectations, no significant correlations appeared between the measures of working/short term memory and the independent variable (Corsi: $r = -.025$, $p = .76$; Backward Digit: $r = .054$, $p = .51$).

Factors Explaining Lexical Diversity Development in the Entire Sample (RQ 1)

To answer our first research question, we fitted backward elimination linear mixed effects models to the data using the lmer() function of the lme4 package for R (Bates et al., 2012). The dependent variable was the Guiraud Index. To control for the influence of household, dyad was included with a random intercept and random slope for time. Time, age group, length of residence, aptitude, working memory, exposure to English, and anxiety when speaking in English were included in the analysis as fixed effects. We computed the variance inflation factor using the vif() function of the car package for R (Fox et al., 2012), and found no problem of collinearity among the predictors (the values ranged between 1.3 and 4.7). Time and age group were treated as factors (T1, T3; adults, children) while the other independent variables were mean-centered and standardized (z_scores). At each step of the analysis, the predictor with the largest nonsignificant p-value was removed until the simplest model was found. At each step, the model with the predictor was compared to the model without the predictor to determine the predictor's significance by using the likelihood ratio test in the anova() function of the lme4 package. The explained variance (marginal and conditional R^2) of the best

Table 2. Mean, standard deviations and paired t-tests of the score on the nine independent variables for each group.

	Group	N	Mean	SD	Min	Max	Paired t-test
LLAMA_B (vocabulary learning)	Adults	38	21.58	10.60	5	50	t(37) = −2.7, p = .01
	Children	38	29.08	13.45	0	55	
LLAMA_D (sound discrimination)	Adults	38	31.93	16.99	0	80	t(37) = −1.27, p = .21
	Children	38	36.84	20.25	0	67	
LLAMA_E (sound-symbol association)	Adults	38	41.84	25.13	0	100	t(37) = .63, p = .53
	Children	38	37.90	27.03	0	100	
LLAMA_F (inductive learning)	Adults	38	14.11	14.84	0	50	t(37) = −2.1, p = .04
	Children	38	20.92	19.78	0	80	
Corsi Block (short-term memory)	Adults	38	4.54	0.95	1	6	t(37) = .36, p = .72
	Children	38	4.46	0.98	2.5	6.5	
Backward Digit (working memory)	Adults	38	4.16	1.48	2	8	t(37) = 1.34, p = .19
	Children	38	3.79	1.30	0	7	
Length of residence	Adults	38	20.66	15.42	0.5	75	t(37) = .93, p = .36
	Children	38	18.55	12.46	0.5	45	
Exposure to English	Adults	38	44.67	13.43	23.06	79.37	t(37) = −9.5, p < .001
	Children	38	66.32	12.37	39.05	91.15	
Anxiety	Adults	38	61.29	14.43	34.44	100	t(37) = 4.78, p < .001
	Children	38	43.60	15.77	20	82	

Table 3. Pearson's correlations matrix between the Guiraud Index (dependent variables) and the nine independent variables to be included in the linear mixed models.

Variable	1	2	3	4	5	6	7	8	9	10
1. Guiraud	—									
2. LLAMA_B	.348***	—								
3. LLAMA_D	.263**	.215*	—							
4. LLAMA_E	.206*	.171*	.028	—						
5. LLAMA_F	.223**	.154	.225**	.159	—					
6. Corsi Span	−.025	.215	.228**	.270**	.061	—				
7. Digit Span	.054	.194	.263**	.198*	.308***	.503***	—			
8. LoR	.295***	−.015	.173*	.028	.160*	−.138	−.252**	—		
9. Exposure	.587***	.356***	.149	.050	.248**	−.019	−.049	.136	—	
10. Anxiety	−.489***	−.172*	−.105	−.008	−.074	.215**	.300***	−.191*	−.638***	—

*p < .05, **p < .01, ***p < .001

Table 4. Best-fitting Model for the Entire Sample

Fixed effects	Best-fitting model		
	Estimate	SE	p
Intercept (adults at Time 1, with all other predictors = 0)	**4.14**	**.18**	**<.001**
Time 3	.18	.15	.24
Group (children)	−.38	.22	.09
Length of residence (z_score)	**.25**	**.11**	**.02**
LLAMA_D (z_score)	**.28**	**.08**	**<.001**
LLAMA_E (z_score)	**.22**	**.08**	**.004**
Exposure to English (z_score)	**.69**	**.12**	**<.001**
Interaction Time 3 by group (children)	**.62**	**.21**	**.005**

fitting model was then computed using the rsquaredGLMM() function of the MuMIn package for R (Barton, 2020).

The results of the best fitting model (marginal R^2 = .40, conditional R^2 = .73) are summarized in Table 4. As the table shows, length of residence, LLAMA_D, LLAMA_E, exposure to English, and the interaction of time by group are significant predictors of lexical diversity. For each increase of one standard deviation in length of residence, the adults score .25 units higher than the adults at T1 (intercept). Also, for each increase of one standard deviation in LLAMA_D and LLAMA_E, their score increases by, respectively, .28 and 22 units. Exposure to English is the most important predictor, with an increase in lexical diversity of .69 units for each increase of one standard deviation in exposure to English. As shown in the model, group and time are *not* significant predictors of lexical diversity development, but the interaction of time by group is: At T3, the children score .42 units higher than the adults at T1.

Factors Explaining Lexical Diversity Development in Adults and Children Separately (RQ 2)

To follow up on the significant interaction between age group and time, in a second step of the analysis, we fitted linear mixed effects models separately to the adults' and children's data. The independent variables were centered and standardized by group (z_score). We followed the same process described above, removing the predictor with the highest p-value at each step and comparing the model with the predictor to the model without the predictor to determine the predictor's significance. We then computed the explained variance of the best fitting models in each group (Adult group: marginal R^2 = .33, conditional R^2 = .76; Child group: marginal R^2 = .54, conditional R^2 = .81).

The results of the best-fitting models for both groups are shown in Tables 5 and 6 below. In the adult group, the only significant predictor of lexical diversity was exposure to English, with a .63 unit increase for each increase of one standard deviation in exposure compared to the adult score at T1. In the child group, time, length of residence, LLAMA_D, LLAMA_E and anxiety are significant predictors of lexical diversity. The most important predictor is time (at T3, children score .80 units higher than at T1), followed by anxiety (decrease of .51 units of lexical diversity for each increase of one

Table 5. Best-fitting Model for the Adult Group.

	Best-fitting model Adults		
Fixed effects	Estimate	SE	p
Intercept (T1, all other predictors = 0)	3.74	.15	<.001
Exposure (z_score)	.63	.13	<.001

Table 6. Best-fitting Model for the Child Group.

	Best-fitting model Children		
Fixed effects	Estimate	SE	p
Intercept (T1, all other predictors = 0)	4.17	.16	<.001
T3	.80	.14	<.001
LoR (z_score)	.44	.16	.01
LLAMA_D (z_score)	.38	.15	.02
LLAMA_E (z_score)	.30	.15	.046
Anxiety (z_score)	−.51	.15	.002

standard deviation in anxiety). Length of residence is also a significant predictor: for each increase of one standard deviation in length of residence, children score .44 score higher. Finally, two dimensions of aptitude have an impact on lexical diversity development in the children: for each increase of one standard deviation in LLAMA_D and LLAMA_E, their scores increase .38 and .30 units, respectively.

Discussion

The LAOC study investigates the effect of age of onset, cognitive variables (aptitude and working memory), exposure to English, and anxiety on English rate of acquisition, measured as the lexical diversity of oral narratives. In the following section, we will review and discuss our results for each dimension.

First, regarding cognitive variables, our results show that two subtests of the LLAMA aptitude battery predict lexical diversity when the entire sample is considered: the higher the participants' scores on the measure of sound discrimination (LLAMA_D) and sound-symbol association (LLAMA_E), the better their lexical diversity development. This effect nevertheless disappears for the adults when modeled separately from the children. This last result contradicts those from the former studies on age of onset and aptitude in naturalistic settings. As a reminder, aptitude showed a predictive effect on ultimate attainment in the adult immigrants in every prior study (Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000; DeKeyser et al., 2010; Granena & Long, 2013a), and in the younger starters in one of the studies (Abrahamsson & Hyltenstam, 2008). Also, in contradiction to former studies using working memory

as a new measure of aptitude, we did not find any effect of working memory on lexical diversity development. This could result either from the measures of working memory themselves, or it could be an indication that working memory predicts language learning in a school context, but not in naturalistic settings, where exposure to the language varies between individuals.

In fact, in our sample, the effect of exposure is of paramount importance for both groups, and greater than the effect of any other variable in the adult group. Following the distinction made by (Moyer, 2011), *quality* of exposure, self-rated in eight domains, seems particularly important in the adults, while length of residence (which Moyer defines as *quantity* of exposure), predicts development of proficiency in the children. In the adult group, exposure to the majority language in everyday life is even more important than length of residence; adults recently arrived but in high contact with the majority language make more progress than adults with a longer stay who do not use English in their everyday lives. This result is surprising considering the range of length of residence present in the data (see Table 2), but it confirms the picture emerging from the 2×2 ANOVAS and paired t-tests, which showed no significant improvement of proficiency in the adult group over a one-year period (see Figure 1).

A closer look at the exposure to English variable shows that children self-report significantly more exposure to English than their parents in all life domains but church (where both children and adults use predominantly Spanish). But it also appears that there is a wide range of exposure within each group (adult group: $M = 45$, $SD = 13$, Min = 23, Max = 79; child group: $M = 66$, $SD = 12$, Min = 39, Max = 91). Adult participants who are in contact with English through work, personal relationships and/or leisure activities significantly increase their chance to develop proficiency in comparison to other participants (including children) who find themselves in almost exclusively Spanish-speaking environments.

Interestingly, exposure to English does not correlate with length of stay in the adult group (Pearson's $r = .08$, $p = .52$), which explains the lack of improvement of lexical diversity in this group over a one-year period. In comparison, in the child group, length of stay correlates with self-rated exposure to English (Pearson's $r = .47$, $p < .001$) and, negatively, with anxiety (Pearson's $r = -.35$, $p = .002$). It appears therefore that the longer children stay in the country, the more contact they have with the language, the less anxious they are when speaking it, and, as a consequence, the more they develop their proficiency.

The study suffers from several limitations. First, the reliability of the LLAMA aptitude tests has been recently questioned (see Bokander and Bylund, 2020), and they might be less adapted to recent immigrants and low-SES participants than to other categories of participants. For another contribution in preparation, we ran ANOVAs comparing the score on the LLAMA tests of the LAOC participants to the scores of the participants from 34 other studies using the same test battery and found out that, except for the LLAMA_D, the adults from the LAOC study performed significantly lower than the adults and teenagers from the other studies, and even lower than the children from the other studies on the LLAMA_B. The children from the LAOC study performed similarly to the children of all the other studies on the four subtests.

Another limitation concerns the dependent variable. As mentioned by an anonymous reviewer of this paper, the Guiraud Index does not consider the various dimensions of lexical knowledge, in particular lexical sophistication (examples such as "the boy looked in the hole of the tree" vs. "the boy gazed into the hollow of the tree", which are identical in terms of lexical diversity, but not in terms of sophistication).

Note however that the Guiraud Index has proved to be a good predictor of untrained raters' perception of the lexical richness of short narratives even in comparison to complex models containing a wide range of lexical properties, including lexical sophistication (see Vanhove et al. 2019). Additionally, it is worth mentioning that the analysis of the other measures of proficiency (verbal tense listening comprehension and verbal fluency) show the same paramount effect of exposure, which appears to be the most stable predictor across all three measures of English proficiency development.

Conclusion

With a sample of 38 parent-child dyads of recently arrived immigrants, this study shows that development of lexical diversity over a one-year period is predicted by exposure (and, for the children, anxiety); foreign language learning aptitude has a smaller effect. We did not replicate results from previous studies showing a differential effect of aptitude as a function of age of onset on ultimate attainment. Contrary to expectations, aptitude measures appeared to be predictors of lexical diversity development in the children but not in the adults. This somewhat surprising result indicates that, at the beginning of adult second language acquisition in a naturalistic setting, exposure to the target language is more important than any other individual factor including working memory and aptitude.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, 30(4), 481–509.

Asher, J. J., & Price, B. S. (1967). The learning strategy of the total physical response: Some age differences. *Child Development*, 38(4), 1219–1227.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic Press.

Barton, K. (2020). *Kamil Barton (2020). MuMIn: Multi-Model Inference. R package version 1.43.17.* Retrieved from https://www.rdocumentation.org/packages/MuMIn/versions/1.43.17.

Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes (2011). R package version* 0.999375-42.

Bokander, L., & Bylund, E. (2020). Probing the internal validity of the LLAMA language aptitude tests. *Language Learning*, 70(1), 11–47.

Carroll, J. B. (1962). The Prediction of Success in Intensive Foreign Language Training. In R. Glaser (Ed.), *Training research and education* (pp. 87–136). University of Pittsburg Press.

Carroll, J. B. (1973). Implications of aptitude test research and psycholinguistic theory for foreign language teaching. *Linguistics*, 11(112), 5–14.

De Clercq, B., & Housen, A. (2019). The development of morphological complexity: A cross-linguistic study of L2 French and English. *Second Language Research*, *35*(1), 71–97.

DeKeyser, R. (1994). How implicit can adult second language learning be. In J. H. Hulstijn & R. Schmidt (Eds.), *Consciousness & Second Language Learning: Perspectives on Form-Focused Instruction, AILA Review, 11*, 83–96.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*(4), 499–533.

DeKeyser, R. (2008). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.) *The Handbook of Second Language Acquisition* (pp. 313–348). Blackwell Publishing.

DeKeyser, R. (2019). The Future of Language Aptitude Research. In Z. E. Wen, P. Skehan, A. Biedroń, & R. L. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (pp. 317–329). Routledge.

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-Linguistic Evidence for the Nature of Age Effects in Second Language Acquisition. *Applied Psycholinguistics*, *31*(03), 413–438. https://doi.org/10.1017/S0142716410000056.

DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & An. M. B. De Groot (Eds.) *Handbook of bilingualism: Psycholinguistic approaches* (pp. 89–108). Oxford University Press.

Dewaele, J.-M., Petrides, K. V., & Furnham, A. (2008). Effects of trait emotional intelligence and sociobiographical variables on communicative anxiety and foreign language anxiety among adult multilinguals: A review and empirical investigation. *Language Learning*, *58*(4), 911–960. https://doi.org/10.1111/j.1467-9922.2008.00482.x.

Dewaele, J.-M., & Sevinç, Y. (2017). La double anxiété langagière des immigrants [The double language anxiety of immigrants]. *Babylonia Journal of Language Education*, *1*, 26–30.

Doughty, C. J., Campbell, S. G., Mislevy, M. A., Bunting, M. F., Bowles, A. R., & Koeth, J. T. (2010). Predicting near-native ability: The factor structure and reliability of Hi-LAB. *Selected proceedings of the 2008 Second Language Research Forum* (pp. 10–31).

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*(2), 141–172.

Ellis, R. (2014). Grammar Teaching for Language Learning. *Babylonia*, *2*, 10–15.

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, *9*(2), 147–171.

Flege, J. E. (1987). A critical period for learning to pronounce foreign languages? *Applied Linguistics*, *8*(2), 162–177.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*(4), 527–552.

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., & Graves, S. (2012). Package 'car.' R Foundation for Statistical Computing, Vienna, Austria.

Garcia de Blakeley, M., Ford, R., & Casey, L. (2017). Second language anxiety among Latino American immigrants in Australia. *International Journal of Bilingual Education and Bilingualism*, *20*(7), 759–772.

Gkonou, C., Daubney, M., & Dewaele, J.-M. (Eds.) (2017). *New insights into language anxiety: Theory, research and educational implications*. Multilingual Matters.

Granena, G. (2014). Language Aptitude and Long-term Achievement in Early Childhood L2 Learners. *Applied Linguistics*, *35*(4), 483–503. https://doi.org/10.1093/applin/amu013.

Granena, G., & Long, M. H. (2013a). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311–343.

Granena, G., & Long, M. H. (2013b). *Sensitive periods, language aptitude, and ultimate L2 attainment* (17618376). John Benjamins Publishing Company.

Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, *63*, 87–106.

Kinsella, C., & Singleton, D. (2014). Much more than age. *Applied Linguistics*, *35*(4), 441–462.

Kiss, C., & Nikolov, M. (2005). Developing, piloting, and validating an instrument to measure young learners' aptitude. *Language Learning*, *55*(1), 99–150.

Krashen, S. D. (1981). Aptitude and attitude in relation to second language acquisition and learning. *Individual differences and universals in language learning aptitude*, 155–175.

Krashen, S. D., Long, M. A., & Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL Quarterly*, 573–582.

Lambelet, A., & Berthele, R. (2015). *Age and foreign language learning in school*. Palgrave Macmillan. http://dx.doi.org/10.1057/9781137525901.

Li, S. (2015). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics*, *36*(3), 385–408.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development*. Springer.

Mayer, M. (1967). *A boy, a dog, and a frog*. Dial Books for Young Readers.

Mayer, M. (1969). *Frog, where are you?* Dial Books for Young Readers.

Meara, P. (2005). *LLAMA, Language Aptitude Tests: The manual*. University of Wales, Swansea. Retrieved from http://www.lognostics.co.uk/tools/llama/llama_manual.pdf.

Moyer, A. (2011). An investigation of experience in L2 phonology: Does quality matter more than quantity? *Canadian Modern Language Review*, *67*(2), 191–216.

Mueller, S. T., & Piper, B. J. (2014). The psychology experiment building language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods*, *222*, 250–259.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, *25*, 46–73.

Robinson, P. (2007). Aptitudes, abilities, contexts, and practice. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from cognitive psychology and applied linguistics* (pp. 256–286). Cambridge University Press.

Robinson, P. (2012). Individual differences, aptitude complexes, SLA processes, and aptitude test development. In M. Pawlak (Ed.), *New perspectives on individual differences in language learning and teaching* (pp. 57–75). Springer.

Ross, S., Yoshinaga, N., & Sasaki, M. (2002). Aptitude-exposure interaction effects on Wh-movement violation detection by pre-and-post-critical period. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 267–300). John Benjamins.

Sáfár, A., & Kormos, J. (2008). Revisiting problems with foreign language aptitude. *IRAL-International Review of Applied Linguistics in Language Teaching*, *46*(2), 113–136.

Sánchez Abchi, V., Bonvin, A., Lambelet, A., & Pestana, C. (2017). Written narrative by Spanish heritage language speakers. *Heritage Language Journal*, *14*(1), 70–93.

Sevinç, Y. (2018). Language anxiety in the immigrant context: Sweaty palms? *International Journal of Bilingualism*, *22*(6), 717–739. https://doi.org/10.1177/1367006917690914.

Sevinç, Y., & Backus, A. (2017). Anxiety, language use and linguistic competence in an immigrant context: A vicious circle? *International Journal of Bilingual Education and Bilingualism*, 1–19. https://doi.org/10.1080/13670050.2017.1306021.

Singleton, D. (2005). The Critical Period Hypothesis: A coat of many colours. *International Review of Applied Linguistics in Language Teaching*, *43*(4), 269–285.

Singleton, D. (2014). Apt to change: The problematic of language awareness and language aptitude in age-related research. *Studies in Second Language Learning and Teaching*, *4*(3), 557–571.

Singleton, D., & Ryan, L. (2004). *Language acquisition: The age factor* (Vol. 47). Multilingual Matters.

Skehan, P. (1991). Individual differences in second language learning. *Studies in Second Language Acquisition*, *13*(2), 275–298.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–94). John Benjamins.

Snow, C. E., & Hoefnagel-Hohle, M. (1978). The Critical Period for Language Acquisition: Evidence from Second Language Learning. *Child Development*, *49*(4), 1114–1128. https://doi.org/10.2307/1128751.

Strömqvist, S. E., & Verhoeven, L. E. (2004). *Relating events in narrative, Volume 2: Typological and contextual perspectives*. Lawrence Erlbaum Associates Publishers.

Treffers-Daller, J. (2009) Code-switching and transfer: An exploration of similarities and differences. In B. E. Bullock & A. J. Toribio (Eds.), *The Cambridge handbook of linguistic code-switching* (pp. 58–74). Cambridge University Press.

UN DESA. (2019, September 17). *The number of international migrants reaches 272 million, continuing an upward trend in all world regions, says UN*. United Nations Department of Economic and Social Affairs (UN DESA). Retrieved from https://www.un.org/development/desa/en/news/population/international-migrant-stock-2019.html.

Vanhove, J., Bonvin, A., Lambelet, A., & Berthele, R. (2019). Predicting human lexical richness ratings of short French, German, and Portuguese texts using text-based indices. *Journal of Writing Research*, *10*(3), 499–525. https://doi.org/10.17239/jowr-2019.10.03.04.

Wen, Z. (2019). Working memory as language aptitude: The phonological/executive model. In Z. E. Wen, P. Skehan, A. Biedroń, & R. L. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (pp. 276–338). Routledge.

Wen, Z., Biedroń, A., & Skehan, P. (2016). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, *50*(1), 1–31.

Wen, Z., & Skehan, P. (2011). A new perspective on foreign language aptitude research: Building and supporting a case for "working memory as language aptitude." *Ilha Do Desterro, A Journal of English Language, Literatures in English and Cultural Studies*, *60*, 15–44.