The Pennsylvania State University

The Graduate School

ACQUIRING GRAMMATICAL CLASSES FROM DISTRIBUTIONAL PROPERTIES OF THE INPUT: THE CHALLENGE OF BILINGUAL ACQUISITION

A Thesis in

Psychology

by

Allison Link

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2021

The thesis of Allison Link was reviewed and approved by the following:

Daniel J. Weiss
Professor of Psychology and Linguistics
Thesis Advisor

Matthew T. Carlson
Associate Professor of Spanish and Linguistics

Elisabeth A. Karuza
Assistant Professor of Psychology

Kristin A. Buss
Professor of Psychology
Head of the Department of Psychology

ABSTRACT

An important development in language acquisition research has been determining the structural properties of language that can be extracted from distributional information in linguistic input. This challenge is compounded when multiple languages are acquired. Recent statistical learning (SL) studies have discovered that learners struggle to acquire multiple languages in the absence of contextual cues. There have also been mixed findings with respect to whether monolinguals and bilinguals differ in their ability to form multiple representations. This study investigated how monolinguals and bilinguals contend with statistical learning of two artificial languages with different grammatical structures, probing the conditions under which learners encapsulate statistics versus generalizing across languages. We created two artificial languages with grammatical classes that could only be derived from distributional cues (following Reeder, 2013). Exp. 1A exposed participants to one of two languages with contrasting phonetic and phonotactic patterns based loosely on Romance and Slavic languages. Monolingual English-speakers learned both languages equally despite their non-native phonetic and phonotactic properties. Exp. 1B investigated whether these languages could be acquired in tandem when they contained conflicting grammatical structures and whether acquisition was dependent on the presence of contextual cues across both monolinguals and bilinguals. We found that learners were able to acquire the underlying grammatical structures of both languages and that learning did not seem to be impacted by language history.

TABLE OF CONTENTS

LIST OF FIGURES v

LIST OF TABLES vi

ACKNOWLEDGEMENTS vii

Chapter 1 Introduction 1

- Grammatical Categorization 1
- Statistical Learning 2
- Bilingual Language Acquisition 5
- Present Study 8

Chapter 2 Experiment 1A 10

- Methods 10
  - Subjects 10
  - Stimulus Materials 10
  - Procedure 13
- Results 15
  - Discussion 17

Chapter 3 Experiment 1B 18

- Methods 18
  - Subjects 18
  - Stimulus Materials 19
  - Procedure 21
- Results 24
  - Learning by Language History 24
  - Learning by Familiarization Structure 27
  - Learning by Familiarization Lexicon 29
  - Reduplicated Strings 30

Chapter 4 General Discussion 31

References 36

Appendix Test Strings 40

LIST OF FIGURES

Figure **2-1**: Raw rating scores for test strings from the Romance and Slavic lexicons..............................16

Figure **3-1**: Raw rating scores for test strings from Parts 1, 2, and 3 by language history.........................26

Figure **3-2**: Raw rating scores for test strings from Parts 1, 2, and 3 by grammatical structure 28

Figure **3-3**: Raw rating scores for test strings from Parts 1, 2, and 3 by lexicon.......................................30

LIST OF TABLES

Table **2-1**: Exp. 1A: Assignment of Words to Categories for the Romance and Slavic Languages 11

Table **2-2**: Exp. 1A: Grammatical AXB Strings Used During Exposure and Test 12

Table **3-1**: Exp. 1B: Assignment of Words to Categories for the Romance and Slavic Languages..........20

Table **3-2**: Exp. 1B: Strings Used During the Adjacent & Non-Adjacent Familiarizations......................21

Table **3-3**: Means and SDs for Ratings during Parts 1, 2, and 3 by Test String and Language History.........25

Table **3-4**: Means and SDs for Ratings during Parts 1, 2, and 3 by Test String and Familiarization Structure.........27

Table **3-5**: Means and SDs for Ratings during Parts 1, 2, and 3 by Test String and Familiarization Lexicon 29

ACKNOWLEDGEMENTS

This thesis could not have been completed without the overwhelming guidance and support I received from my fellow lab mates, Amy Lebkuecher and Natalie Awad Schwob. At every challenge I have endured since enrolling in my graduate studies, these two have been by my side to provide me with unconditional encouragement and assistance. Every paper, every talk, every dataset, and every analysis were influenced and improved upon by their expertise and advice. I can say for certain that I would not be where I am today if not for their mentorship. I wish you two nothing but the best as you complete your dissertations, and I cannot wait to watch you thrive in your future endeavors!

Chapter 1

Introduction

Grammatical Categorization

One of the well-known challenges associated with acquiring a productive linguistic system is the impoverished nature of the input (Chomsky, 1965). Learners do not experience every possible context in which a given word can be used and nevertheless must figure out how correctly extend usage. Forming grammatical categories (e.g., nouns, verbs, adjectives, etc.) is one of the tools used to advance language acquisition despite the limited number of linguistic contexts experienced by learners. The process of grammatical categorization allows learners to organize new words into categories based on characteristics shared by category members. Grammatical categorization permits productivity as it allows learners to generalize based on previously acquired information and thereby determine which contexts are appropriate for a newly encountered word. The ability to identify words from speech input and categorize them into grammatical classes is an important skill that precedes the formation of organized speech, as understanding grammatical classes is necessary before a word can be used properly in a sentence (Labelle, 2005). The challenge of determining category membership by generalizing properties of previously experienced words to new words is compounded by word forms being able to belong to more than one category (e.g., the word *run* can be a noun or verb) and the inconsistent relationship between functional, semantic, and perceptual features among members of grammatical classes (Labelle, 2005).

The question of how learners can successfully use a new word or a learned word in a new context without violating grammatical rules has evoked several hypotheses regarding the nature and formation of grammatical categories (Gleitman & Wanner, 1982). A prominent view of grammatical categories is that they are innately specified (e.g., Chomsky, 1965), available for children before ever receiving language input. Universal part-of-speech categories have been criticized as they do not sufficiently explain the considerable cross-linguistic variability in how grammatical classes and their subclasses are instantiated

(Culicover, 1999). For example, some languages divide pronouns further based on gender and some do not. Another popular theory regarding the formation of grammatical classes is semantic bootstrapping, which builds on the assumption of innate grammatical categories by proposing a mechanism by which lexical items in a child's environment might be assigned to categories (Grimshaw, 1981). This theory posits that in the early stages of language development, children may analyze the semantic properties of words and assign words with similar meaning or function to the same predefined grammatical categories (e.g., Pinker, 1984, 1987). For example, learners might identify words that denote action as verbs. This is complicated by the fact that the semantic properties of words do not always map reliably onto grammatical categories (Maratsos & Chalkley, 1980), such as verbs that do not describe a physical action (e.g., *to want*). This suggests that learners must draw on additional sources of information in order to place words into grammatical categories. In sum, there is a lack of consensus with respect to how learners form grammatical classes during language acquisition. Theories regarding the nature of grammatical categorization must take into account that words can belong to multiple grammatical classes, and that members of a given category do not typically share consistent functional, semantic, or perceptual features. As such, researchers have considered that learners may look beyond the meaning of a word and also examine information embedded within the surrounding context in order to make category judgements (Maratsos & Chalkley, 1980).

Statistical Learning

In the context of grammatical categories, *distributional learning* refers to the process by which learners detect regularities in the linguistic environment of a word and use them utilize to deduce category structure (Maratsos & Chalkley, 1980). Words that typically occur in the same contexts could be placed into the same grammatical category. For example, a learner might identify that English word roots ending in /-ed/ can also end in /-s/ and subsequently consider these words to be verbs (Maratsos & Chalkley, 1980; see Redington et al., 1998). A substantial body of research has shown that distributional learning (also referred to as *statistical learning* or *SL*) may play a role nearly all areas of language acquisition. A landmark study by Saffran and colleagues (1996) showed that 8-month-old infants were sensitive to distributional information embedded in artificial speech and could use it to identify word boundaries. While research on acquiring regularities in the context of artificial grammar learning had existed well before this study (e.g., Reber, 1967), the implications for its findings on the role of distributional learning in early language acquisition had a significant impact on the field. Since this discovery, researchers have found evidence for the role of distributional information in the learning of phonetic and phonotactic properties of language (Maye et al., 2002; Maye et al., 2008, Chambers et al., 2003) as well as word learning (e.g., Yu & Smith, 2007).

A longstanding premise suggested that learners were unable to rely on distributional information in order to form linguistic categories (Braine, 1966). It seemed unlikely that learners could navigate the variability inherent in linguistic input while successfully extracting distributional information for category formation. Further, learners would have to deduce whether a lack of information is meaningful or simply implies it has yet to be encountered in the input (Braine, 1987; see Reeder et al., 2013). Researchers have argued that while learners may not be able to utilize distributional cues exclusively when forming categories, they may be successful when these cues are coupled with other relevant features in the input. Correlated cues are features of the input that are typically linguistic or perceptual in nature and may aid learners in identifying underlying linguistic structure (e.g., Braine, 1966; Reeder et al., 2013). Linguistic correlated cues can be phonological, morphological, or semantic in nature (e.g., words ending /-ist/ refer to people of a certain profession and are nouns) while perceptual correlated cues (e.g., the repetition of items in a sequence; Gomez & Gerken, 1999) are auditory features of the input that are nonlinguistic in nature.

The role of distributional information in the formation of grammatical categories is of particular interest. While the presence of correlated cues can help guide learners in their categorizations, there are not typically one-to-one mappings between a word's phonological, morphological, or semantic features and the grammatical category to which they belong (Gleitman, 1990). Reeder and colleagues (2013) conducted a study to examine whether learners could utilize different kinds of distributional information in the absence of correlated cues in order to deduce category structure. During their first experiment, monolingual English-speaking subjects listened to a stream of sentences in an artificial language. The language was generated by a grammatical structure wherein each sentence contained several categories of nonsense words. The language did not contain any correlated cues that could indicate its structure, so subjects could only acquire it through the underlying distributional properties embedded within the grammar. Participants heard only a subset of possible grammatical category combinations during exposure and the remaining combinations were presented during test to see if subjects could generalize the grammatical structure that they heard to new contexts. Importantly, subjects were exposed to speech that was linguistically and perceptually similar to their native language as the nonsense words were based on English phonology and phonotactics and were recorded by a native English speaker. They found that subjects were in fact able to learn the structure of the language using only distributional cues and that they were able to generalize the structure to novel contexts. Learners are also systematic in how they utilize distributional information: stronger distributional cues (marked by less contextual overlap and increased exposure) allow for generalization to new words that share very few contexts with other category members, while weaker cues (marked by more contextual overlap and decreased exposure) result in fewer generalizations based on distributional information.

It is important to note that the logic employed by Reeder and colleagues (2013) is based on exposure to a single language. If one considers that the majority of the world's population is bilingual (Crystal, 2011), the challenge of grammatical categorization for many learners may be deducing the categories as they are instantiated across multiple languages. This poses the question of whether bilingual learners can still succeed in acquiring grammatical categories across two languages without the aid of correlated cues.

Bilingual Language Acquisition

Following the pivotal study by Saffran and colleagues, SL researchers began to explore the extent to which learners of all ages could identify and track distributional information across linguistic input (see Frost et al., 2019 for review). Overwhelmingly, SL research has adopted the perspective of monolingual speakers by examining learning during exposure to a single statistical pattern that is invariant across the familiarization period (see Bulgarelli et al., 2018 for review). This approach does not consider that young learners encounter significant variability and noise in their linguistic environments. For bilingual learners, this challenge is compounded by the fact that multiple languages are interleaved in the input. The question of how bilingual learners track the wide variety of distributional regularities in their input to acquire language is the focus of a subset of research within the SL community.

Recent research in bilingual SL has approached bilingualism in two ways (see Bulgarelli et al., 2018 for review). One perspective has been to investigate how learners cope when multiple patterns are present in the input, such as two artificial languages presented sequentially. The conclusion from this line of research has been that learners, under some circumstances, can acquire two patterns sequentially, though it is generally difficult without the aid of additional correlated cues that help to distinguish either language. A second approach has explored whether monolingual and bilingual speakers differ when acquiring either one or multiple patterns. There have been rather mixed findings thus far, with little consensus on when and why bilingual learners track statistics of their input differently relative to monolinguals (see Bulgarelli et al., 2018 for review).

One of the first studies that explored how subjects contend with multiple patterns in the input was conducted by Weiss and colleagues (2009). They examined whether learners could form separate, encapsulated statistical representations for two different artificial languages. Across several experiments, subjects were passively exposed to two streams of artificial speech presented in an interleaved fashion. The distributional information embedded within the two streams conformed to one of two conditions. When the statistics were congruent, learners should have been able to segment both streams regardless of whether they formed separate or combined statistical representations. When the statistics were incongruent, learners would be most successful in segmenting the streams when they encapsulated the statistics from each language because their different statistical structures made it difficult to form combined representations. They found that when learners were exposed to two different artificial speech streams spoken in different voices (female and male), they succeeded in segmenting the streams in both the congruent and incongruent conditions. When the cue of speaker voice was stripped away such that both streams were presented with the same voice, only participants in the congruent condition were able to segment the streams. These findings suggest that learners are able to form different statistical representations from two artificial languages with the aid of perceptual correlated cues. Gebhart and colleagues (2009) conducted another study aimed at investigating how learners navigate multiple patterns in the input. Specifically, they looked at the conditions under which subjects listening to a speech stream could identify a change in statistical structure and subsequently learn both sets of statistics. Subjects listened to a stream of synthesized speech in which the underlying distributional pattern at the beginning was changed mid-stream to a different pattern. They found that when subjects were not given any explicit instruction to this change or any additional correlated cues, learning only occurred for the first structure. When subjects experienced a shift in pitch during the structural change such that the languages sounded slightly different, they were still unable to learn the second structure. When subjects were told that they would be listening to two different languages (without a change in pitch) and that the mid-stream change would be marked by a long pause, participants were able to learn both structures equally well. While changing the phonological qualities of the two languages did not aid participants in detecting a structural change, an explicit cue to this change was sufficient for learners to acquire both structures. Taken together, both of these studies suggest that learners are capable of acquiring multiple statistical patterns. They also emphasize the significant role that correlated cues play in the successful acquisition of distributional patterns during SL tasks, specifically in the case of multiple patterns.

The question of whether monolingual and bilingual speakers track distributional information differently is a topic of current debate. Poepsel and Weiss (2016) found evidence that bilingual speakers may acquire distributional mappings more efficiently than monolinguals, suggesting that language history impacts statistical learning. Subjects completed a cross-situational statistical learning task where during familiarization, they observed novel objects appear on a screen while listening to nonsense words. Some trials contained 1:1 object-word mappings in which one word was only heard in the presence of one specific object, while other trials contained 2:1 mappings where one word was heard alongside two different objects. Afterwards, subjects were tested on their object-word mappings. They were presented with a single nonsense word alongside two novel objects (only one of which was paired with the nonsense word during familiarization) and asked to select which word best described the object. They found that while both monolingual and bilingual subjects were able to learn the 1:1 mappings equally well, the bilingual participants acquired more 2:1 mappings than monolinguals and did so during the earlier test phases. This suggests that bilinguals were more accepting of the possibility of multiple mappings in the input and that they could form these mappings with less exposure than monolinguals. Contrary to these findings, there is also evidence to support that bilingualism may not impact how learners contend with statistical information. Bulgarelli et al., (2019) examined whether bilingual speakers would perform differently than monolingual speakers on a visual statistical learning task. Participants viewed a stream of shapes that changed orientation on the screen, controlling for the features and location of each shape. Afterwards, subjects were tested on whether they learned the co-occurrence probabilities between the location of the shapes and their features. They found that both monolingual and bilingual participants were able to track both sets of statistical information equally well in spite of their different language backgrounds. The contrasting results from these two studies highlight the lack of consensus regarding how speakers' statistical representations are impacted by their language history.

Many studies of bilingual SL, including those described above (e.g., Weiss, Gerfen, and Mitchel, 2009), have embraced the notion that successful learning is contingent on maintaining completely distinct representations for each language heard during exposure. However, a fundamental insight of bilingualism research is that bilinguals do not function as two individual monolinguals (Grosjean, 2008). There is copious evidence to suggest that there is a bidirectional cross-linguistic influence of bilinguals' first and second languages on their linguistic representations (e.g., Serratrice, 2013). This influence may also extend to statistical learning, particularly in whether learners' representations of statistical regularities change as a result of exposure to another language.

Present Study

The present study sought to examine how monolingual and bilingual learners contend with the distributional learning of grammatical categories when there are two artificial languages present in the input. This series of experiments was based heavily on the work by Reeder and colleagues' (2013). Our goal was to replicate and expand on the original findings of their first experiment. In our study, Experiment 1A aimed to examine whether the findings from the original experiment could be replicated when subjects were exposed to an artificial language with a non-native phonology and phonotactics. Our participants were exposed to the same sentences generated by the same grammar as those of Reeder and colleagues, with the only difference being the lexical items assigned to each grammatical category. Subjects were assigned to one of two conditions: the "Romance" lexicon contained nonsense words based on words in Spanish and the "Slavic" lexicon contained nonsense words based on words in Serbian. Both sets of stimuli were recorded by a female speaker who is fluent in both Spanish and Croatian and spoke using an appropriate accent for either lexicon. Our goal was to examine whether the results of the original experiment were replicable despite these methodological changes, paving the way for subsequent experiments which would introduce multiple languages during exposure. Additionally, we recruited our subjects from Prolific.com and ran this task entirely over the internet in order observe the feasibility of online recruitment methods which could aid in recruiting specific bilingual populations as we planned to do in Experiment 1B.

Assuming participants learn the underlying grammatical structure presented in Exp. 1A, Exp. 1B examined whether participants could acquire two different artificial grammars presented sequentially. Participants were presented with two languages during familiarization that each utilized a different modified version of the grammar from Exp. 1A. One language was presented with the Romance lexicon and the other was presented with the Slavic lexicon, which we anticipated would cue subjects to their differing grammatical structures and aid in learning. We tested both monolingual English speakers and English-speaking bilinguals from Prolific.com to observe whether language history impacted learning.

Chapter 2

Experiment 1A

The goal of this experiment was to replicate the results of Reeder and colleagues' (2013) first experiment with two languages in order to determine whether monolingual English-speaking participants could acquire grammatical categories that contained non-native phonological and phonotactic properties.

Methods

Subjects

163 monolingual native English-speakers between the ages of 18 – 30 years old with no history of language or hearing disorders were recruited from Prolific.com. Of these participants, there were 43 who were excluded from our analyses. We excluded participants that failed multiple attention checks, failed the headphone check, and/or did not follow the instructions (20) and those who selected the same rating for every item during test (23). The remaining 120 participants (50 female, 70 male) were included in the analyses. Subjects were compensated $5.55 for completing the study. Participants were pseudorandomly assigned to one of four language conditions, Romance A ($N = 30$), Romance B ($N = 30$), Slavic A ($N = 30$), or Slavic B ($N = 30$).

Stimulus Materials

We created two artificial languages by generating two different pseudoword lexicons and assigning lexical items to the grammatical structure outlined in the first experiment of Reeder and colleagues (2013). Both languages had the same number of lexical items and followed the same grammatical structure, such that the only difference between the two were their lexical inventories. The goal was to create two languages that would be audibly distinct to listeners by incorporating phonological and phonotactic differences in the lexical inventory and by incorporating different speaker accents during recording.

The first lexicon, referred to as the Romance lexicon, was comprised of pseudowords based on Spanish. To create the lexicon, we utilized Wuggy (Keuleers & Brysbaert, 2010), an application that generates pseudowords based on an inventory of real words that users can input. Through the statistical program R (R Core Team, 2015), we utilized the *vwr* package (Keuleers, 2015) which contains several hundred thousand words from eight different languages. We selected a list of 114 Spanish words (out of 31,490 available) that varied from one to two syllables in length and counterbalanced the onsets and offsets such that there were equal proportions of consonants and vowels in those positions. We then imported this list into Wuggy and selected thirteen of the generated pseudowords to comprise the Romance lexicon (see Table **2-1**). The lexicon was balanced such that there was a relatively equal number of one and two syllable words, and that vowels and consonants were evenly distributed as the onset and offsets of words. The second lexicon, referred to as the Slavic lexicon, was created using the same procedure with an input of Serbian words that were selected from the same R package. The Slavic lexicon also contained thirteen pseudowords, several of which were three syllable words to further distinguish it from the Romance lexicon (see Table **2-1**).

Table **2-1**: Exp. 1A: Assignment of Words to Categories for the Romance and Slavic Languages

(Q)	A	X	B	(R)
Romance A				
gres	pel	bues	hol	cos
onga	ibla	cle	mie	amble
	plamo	urnam	enclim	
Romance B				
cle	bues	pel	cos	gres
urnam	hol	mie	amble	onga
	enclim	ibla	plamo	

(Q)	A	X	B	(R)
Slavic A				
mors	jofar	Debiz	ifort	krur
avre	untrom	skoshi	drege	očira
	ogiru	rnveta	huvanda	
Slavic B				
krur	avre	untrom	jofar	mors
huvanda	skoshi	Drege	debiz	ifort
	očira	rnveta	ogiru	

The grammatical structure used to generate sentences for this experiment was identical to that used in Experiment 1 reported by Reeder and colleagues (2013). The grammar followed the form (Q)AXB(R), where each letter represented a different category of pseudowords. The A, X, and B categories appeared in every sentence while Q and R were flanker categories that were optional. This prevented participants from using word position in a sentence (e.g., initial position, final position) as a cue to the underlying grammatical structure. The Q and R categories each contained two words and the A, X, and B categories each contained three words. Sentences were generated by selecting one word from each category, creating four possible sentence combinations that varied in length from three to five words (AXB, QAXB, AXBR, or QAXBR). There were 27 possible strings that could be generated using only the A, X, and B categories of the grammar (3 A-words x 3 B-words x 3 C-words). 18 of these strings were selected for familiarization such that subjects heard every X-word with every A-word and with every B-word while not exhausting every A-B context (see Table **2-2**). These 18 strings were presented with every possible flanker category combination, resulting in 72 (Q)AXB(R) sentences heard during the familiarization phase. We presented the nine withheld strings during test (see Table **2-2**) to examine whether subjects could generalize the distributional information presented during familiarization to new A-B contexts.

Table **2-2**: Exp. 1A: Grammatical AXB Strings Used During Exposure and Test

A_1	X_1	B_1	A_1	X_2	B_1*	A_1	X_3	B_1
A_1	X_1	B_2*	A_1	X_2	B_2	A_1	X_3	B_2
A_1	X_1	B_3	A_1	X_2	B_3	A_1	X_3	B_3*
A_2	X_1	B_1*	A_2	X_2	B_1	A_2	X_3	B_1
A_2	X_1	B_2	A_2	X_2	B_2	A_2	X_3	B_2*
A_2	X_1	B_3	A_2	X_2	B_3*	A_2	X_3	B_3
A_3	X_1	B_1	A_3	X_2	B_1	A_3	X_3	B_1*
A_3	X_1	B_2	A_3	X_2	B_2*	A_3	X_3	B_2
A_3	X_1	B_3*	A_3	X_2	B_3	A_3	X_3	B_3

Note: ***** denotes strings reserved as Grammatical Novel test items

To create the two artificial languages, we applied the words from the Romance and Slavic lexicons to the (Q)AXB(R) grammar (see Table **2-1**). Each word was only assigned to a single category, and the assignment of words to categories was balanced such that each category had a relatively equal number of one-, two-, or three-syllable words and vowels and consonants were evenly distributed as the onsets and offsets of category members. There were two different versions of the Romance Language, with the only difference being the category assignment of individual words from version A to version B. Similarly, there are two versions of the Slavic language. This controlled for possible idiosyncratic effects arising from particular items or positions.

A female bilingual speaker whose native language is Croatian and who is also fluent in Spanish (and began learning at age six) recorded the stimuli used throughout this study. She was instructed to use a Spanish accent for recording the Romance stimuli and a Croatian accent for recording the Slavic stimuli. Each of the pseudowords from both languages were recorded in Praat (Boersma & Weenink al., 2021) in both a terminal and non-terminal intonation and were adjusted so that the pitch, volume, and syllable duration were relatively consistent. Sentences were created by selecting one word from each category and splicing them together into sequences using Python with 50ms of silence between words and using the terminal intonation recording as the final word in a sentence.

Procedure

Participants registered on Prolific.com who met our inclusion criteria (see Subjects) were eligible to participate in this experiment. They used their personal computer and were informed that headphones would be required for the experiment. After agreeing to participate, participants were directed to Pavlovia.com where they completed the experiment using their internet browser. Before starting the familiarization phase, participants completed a headphone check (derived from Woods et al., 2017) to assess whether they were equipped with a stereo headset. This was a 3-AFC task in which participants heard three pure tones played in succession and were then asked to select which tone was the quietest. One of these tones was in antiphase across the stereo channels, resulting in a softer perception of that tone when played over speakers relative to stereo headphones. Individuals listening through speakers and those wearing headphones would thus consistently select different tones as being the quietest. According to Woods et al. (2017), a minimum of five out of six correct sets is sufficient to judge whether a participant is wearing headphones. Subjects listened to and judged six sets of tones before receiving feedback on whether they passed and were able to proceed with the experiment or failed and were reminded to wear headphones to complete the task once more. Participants that failed both times were excluded from our analyses but continued with the experiment as Prolific does not allow researchers to prevent participants from completing an experiment based on screening information collected after sign-up.

In the instructions, participants were informed that they would listen to sentences from a foreign language that they had never heard before and would later be tested on their memory of those sentences. They were instructed to remain seated at their computer with headphones on for the duration of the experiment and were reminded not to walk away from their computer or navigate to another tab in their browser. In order to ensure that these instructions were followed, frequent attention checks were implemented in between sentences during familiarization. Participants would hear between three and five sentences before encountering an attention check in which one, two, or three tones were played in succession. Both the number of sentences between attention checks and the number of tones played during attention checks were randomly generated. Participants were asked to judge how many tones they heard by selecting a number from one to three, after which the experimental stimuli resumed playing. These attention checks were timed such that the sentences would continue to play if an answer had not been selected after 20 seconds. Participants practiced these attention checks for two trials before beginning familiarization and were informed that they may not receive payment if they failed excessive attention checks (i.e., selected incorrect judgements or failed to answer within 20 seconds for at least 10% of all attention checks). During familiarization, the set of 72 unique sentences was presented randomly without replacement four times with an interstimulus interval of 1.5s of silence between each sentence. This created an exposure set of 288 sentences that varied between 20 to 25 minutes depending on response time to the attention checks.

Following familiarization, participants completed a testing phase in which they judged the grammaticality of 36 AXB test strings (i.e., sentences without flanker categories; see Appendix). There were three types of test strings: grammatical familiar (9 AXB strings that were heard during familiarization), grammatical novel (9 AXB strings that were withheld from familiarization but conformed to the grammar), and ungrammatical (18 strings that did not conform to familiarization; either AXA or BXB). None of the ungrammatical strings contained any reduplicated lexical items. Participants were randomly presented with a single test string and then were asked to rate it on a scale of 1-5 based on whether they thought it came from the language that they heard during familiarization (the same criteria used by Reeder et al., 2013). On every response trial, participants were shown the following key below the rating scale: 1 = Definitely did NOT come from the language; 2 = Might not have come from the language; 3 = May or may not have come from the language; 4 = Might have come from the language; 5 = DEFINITELY came from the language. To provide additional clarification, participants were instructed to go with their gut reaction as to whether the string might have been something a native speaker of the language would have said when following the rules of the language's grammar. After reviewing data from our initial thirty participants, we found that many subjects rated all 36 test strings as a 5. To mitigate this, participants were encouraged to consider the entire rating scale when making a judgement as well as explicitly instructed not to give every single sentence the same rating.

Results

All analyses were conducted using R version 3.6.3 (R Core Team, 2016) with the lmerTest package (Kuznetsova, Brockhoff, & Christiansen, 2017) and orthogonal contrast codes for all categorical variables. For all participants from Exp. 1A (N = 120), we evaluated the main effect of grammar on rating during the grammaticality judgement task by conducting a linear mixed-effects model with the interaction between grammar condition (familiar, novel, and ungrammatical) and lexicon at exposure (Romance or Slavic) as a within-subjects fixed effect and language version (A or B) as a between-subjects fixed effect.

For random effects, we added a by-participant random intercept and by-participant random slope for grammar condition and a by-item random intercept. There was a significant main effect of grammar on subjects' grammaticality ratings (β = .69, t(98.28) = 9.89, p < .001) and no significant interaction between the grammar condition of test strings and the lexicon heard during exposure (β = -.18, t(118) = -1.41, p = .16). We did not find a main effect of exposure lexicon (β = -.07, t(118) = -.78, p = .44) or language version (β = -.02, t(117) = -.21, p = .84) on test ratings. Across both languages, the mean rating of grammatical familiar strings was 3.96 (*SE* = .05), the mean rating of grammatical novel strings was 3.92 (*SE* = .05), and the mean rating of ungrammatical strings was 3.25 (*SE* = .06) (see Figure **2-1** for mean ratings within the Romance and Slavic lexicons).

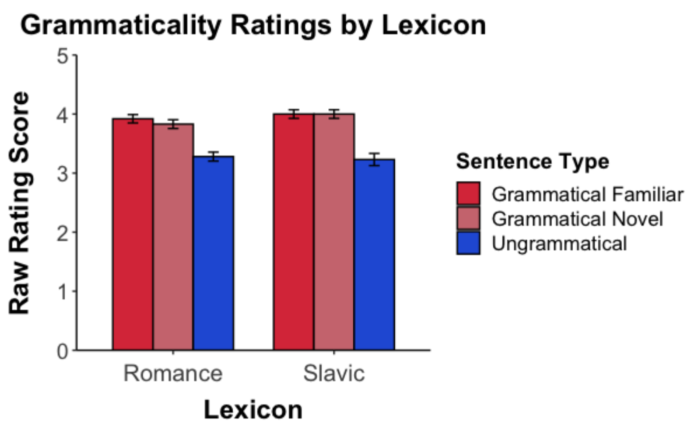

Figure **2-1**: Raw rating scores for test strings from the Romance and Slavic lexicons

Using the same linear mixed-effects model, we modified the variable for grammar condition using orthogonal contrast codes to determine if there were significant differences between ratings of the three different types of test strings. We confirmed that there was no significant difference between participants' ratings of grammatical familiar and grammatical novel strings (β = .04, t(16.44) = .31, p = .76). When comparing the ratings of the two grammatical string types to ungrammatical strings individually, we found that both the grammatical familiar strings (β = .67, t(62.79) = 7.44, p < .001) and the grammatical novel strings (β = .64, t(58.35) = 6.62, p < .001) were rated significantly higher than ungrammatical strings.

Discussion

The results of this study replicated and extended the findings from the first experiment of Reeder et al. (2013). Participants were able to learn the grammatical structure of an artificial language despite having non-native phonology and phonotactics. Monolingual English speakers rated grammatical strings, whether familiar or novel, as more likely to arise from the language they listened to during familiarization relative to ungrammatical sentences. Participants who were exposed to either the A or B versions of the Romance and Slavic languages did not rate sentences differently during test. Overall, there was no significant difference in rating depending on the lexicon of the language at exposure. Taken together, these results indicate that the non-native properties of the Romance and Slavic lexicons do not impact whether English-speaking monolinguals can acquire the underlying structure of an artificial language.

Unlike the original study, these results were found in the context of an online platform. Collecting data over the internet allows us to sample from a wide range of speakers from across the globe and control for specific features of our participants' language history. Given the findings of Exp. 1A, we conducted Exp. 1B in order to examine whether participants could acquire the grammatical structure of two artificial languages with similar properties in the same experimental session. As previously noted, acquisition of multiple patterns can be aided by introducing correlated cues to distinguish either pattern (e.g., speaker voice; Weiss et al., 2009). Learners may make use of phonotactic differences between auditory stimuli in particular to aid in the acquisition of distributional information (see Dal Ben et al., 2021). Thus, in order to facilitate learning of both languages, we provided the phonological and phonotactic differences between the Romance and Slavic lexicons to serve as indexical cues to each language. Since no difference was found between the A and B versions of either lexicon, we opted to use only version B during Exp. 1B.

Chapter 3

Experiment 1B

The first goal of this experiment was to determine whether subjects could deduce the category structure of two artificial languages when they were presented sequentially and contained different grammatical categories. In Exp. 1A, all subjects listened to sentences following the grammatical structure (Q)AXB(R), where A, X, and B each represented a different category of words. In the following experiment, subjects were exposed to one language with the grammar (Q)AAB(R) and another with the grammar (Q)ABA(R). A and B were two distinct categories of words and the A-category was heard twice during each sentence. The primary difference between these two structures is the position of the second A-category (A_2), which was either adjacent (i.e., AAB) or non-adjacent (i.e., ABA) to the first A-category (A_1). These changes were motivated by differences in sentence structure observed in natural languages, such as the distinction between SOV and SVO word ordering. The two languages adopted the Romance and the Slavic lexicons from Exp. 1A. The differing lexical items and speaker accents might serve as correlated cues that could aid learners in distinguishing the languages from each other. Our second goal was to compare performance on the task between monolingual and bilingual participants in order to determine if language history impacts learning, as previous research has produced mixed results as to whether bilingual subjects confer an advantage over monolinguals when faced with learning multiple patterns (see Bulgarelli et al., 2018 for review).

Methods

Subjects

167 participants between the ages of 18 – 30 years who reported no history of language or hearing disorders were recruited from Prolific.com. Of these participants, there were 47 who were

excluded from analyses. We excluded participants that failed multiple attention checks and/or did not follow the instructions (34) and those that failed the headphone check (13). The remaining 120 participants (66 female, 52 male, 2 prefer not to say) were included in the analysis. Participants were compensated an average of $6.89 for completing the study. The monolingual speaker condition (N = 60) was composed of monolingual English-speakers and the bilingual speaker condition (N = 60) contained bilingual speakers who were fluent in English and one additional language that is not phonologically related to Spanish or Serbian. Bilingual subjects were filtered using Prolific's custom prescreen such that speakers of Basque, Bulgarian, Catalan, Croatian, French, Galician, Italian, Macedonian, Portuguese, Serbian, Slovenian, and Spanish could not participate in our study. The bilinguals who were included in our analyses spoke a wide range of languages, including Afrikaans, Arabic, Armenian, Bengali, Chinese, Gaelic, Greek, Polish, Tagalog-Filipino, Turkish, Urdu, and Vietnamese. Participants were randomly assigned to one of four language conditions, Romance ABA → Slavic AAB (Monolingual: N = 15; Bilingual: N = 15), Romance AAB → Slavic ABA (Monolingual: N = 15; Bilingual: N = 15), Slavic ABA → Romance AAB (Monolingual: N = 15; Bilingual: N = 15), or Slavic AAB → Romance ABA (Monolingual: N = 15; Bilingual: N = 15).

Stimulus Materials

Participants were presented with two languages during familiarization, each generated using one of two different grammatical structures. One language had an adjacent structure that could be captured formally as (Q)AAB(R) while the other language had a non-adjacent structure using (Q)ABA(R) (see Table **3-1**). Just as in the previous experiment, each letter of the grammar represents a different category of words. The grammatical structures from the present experiment contain only two unique word categories (A and B), whereas Exp. 1A contained three unique categories (A, X, and B). Both of the grammatical structures in Exp. 1B contained two A-categories and one B-category, with the critical difference being the position of A_2 (either adjacent or nonadjacent to A_1). As in the previous experiment, the A and B categories were present in every sentence while the Q and R categories were optional. Since our analyses from Exp. 1A revealed that there was no difference in learning between the A and B versions of either language, we opted to use only version B of the Romance and Slavic lexicons when assigning words to categories for this experiment (see Table **3-1**). Sentences were generated by selecting one word from each category, creating four possible sentence combinations that varied in length from three to five words for each the adjacent language (AAB, QAAB, AABR, or QAABR) and the non-adjacent language (ABA, QABA, ABAR, QABAR). Sentences in each language were generated using the same methods and auditory stimuli as in Exp. 1A. The ordering of the grammars was counterbalanced such that half of our participants (*N* = 60) heard the adjacent language first and half of our participants (*N* = 60) heard the non-adjacent language first. The lexicons applied to each language were also counterbalanced, in that half of our participants (*N* = 60) heard the Romance lexicon with the adjacent language and the Slavic lexicon with the non-adjacent language while the other half (*N* = 60) heard the opposite configuration. This resulted in a total of four experimental conditions with *N* = 30 in each.

Table **3-1**: Exp. 1B: Assignment of Words to Categories for the Romance and Slavic Languages

(Q)	A	A	B	(R)	(Q)	A	B	A	(R)
Romance (adjacent)					*Romance (non-adjacent)*				
cle	bues	bues	cos	gres	cle	bues	cos	bues	gres
urnam	hol	hol	amble	onga	urnam	hol	amble	hol	onga
	enclim	enclim	plamo			enclim	plamo	enclim	
Slavic (adjacent)					*Slavic (non-adjacent)*				
krur	avre	avre	jofar	mors	krur	avre	jofar	avre	mors
huvanda	skoshi	skoshi	debiz	ifort	huvanda	skoshi	debiz	skoshi	ifort
	očira	očira	ogiru			očira	ogiru	očira	

There were 27 possible ABA or AAB strings that could be generated for either language. The adjacent strings were identical to the non-adjacent strings except the ordering of the B-category and A_2 were swapped. Twelve strings were selected to be used during familiarization (see Table **3-2)**. These strings had complete distributional overlap between the B-category and each of the two A-categories such that subjects heard every A_1-B context and every A_2-B, but not every A_1-A_2 context. Additionally, none of these sentences contained the same lexical item in both A_1 and A_2. Each of the 12 strings were presented with all four combinations of the optional flanker categories (i.e., no flankers, Q only, R only, both Q and R), resulting in 48 (Q)ABA(R) or (Q)AAB(R) sentences presented during familiarization. Both familiarization phases repeated these 48 sentences three times in a pseudorandom order for an exposure set totaling 144 sentences per language.

Six of the 27 strings were withheld from familiarization to be used as grammatical novel stimuli during test to examine whether subjects could generalize the distributional information presented during familiarization to new A_1-A_2 contexts (see Appendix). The nine remaining strings contained reduplication such that both A-categories used the same lexical token. Six of these strings were reserved for the final test (see Appendix) while the remaining three were not used in the experiment. Overall, the proportion of distributional information heard in the familiarization period for a single language during Exp. 1B was approximately one-third of that heard during Exp. 1A.

Table **3-2**: Exp. 1B: Strings Used During the Adjacent & Non-Adjacent Familiarizations

Nonadjacent								
A_1	B_1	A_2	A_1	B_2	A_2	A_1	B_3	A_2
A_1	B_1	A_3	A_1	B_2	A_3	A_2	B_3	A_1
A_2	B_1	A_1	A_2	B_2	A_3	A_2	B_3	A_3
A_3	B_1	A_2	A_3	B_2	A_1	A_3	B_3	A_1
Adjacent								
A_1	A_2	B_1	A_1	A_2	B_2	A_1	A_2	B_3
A_1	A_3	B_1	A_1	A_3	B_2	A_2	A_1	B_3
A_2	A_1	B_1	A_2	A_3	B_2	A_2	A_3	B_3
A_3	A_2	B_1	A_3	A_1	B_2	A_3	A_1	B_3

Procedure

After agreeing to participate on Prolific, subjects were directed to Pavlovia.com where they completed the experiment using an internet browser. Before starting the first familiarization phase, participants completed the same headphone check from Exp. 1A (derived from Woods et al., 2017) to assess whether they were equipped with a stereo headset. Subsequently, participants were informed that the experiment contained three parts. In Part 1, they would listen to sentences from a foreign language that they had never heard before, followed by a memory test of those sentences. In Part 2 they would repeat the same procedure as in Part 1, and in Part 3 they would complete a final test. There were no instructions which explicitly stated that subjects would be listening to two different languages.

The procedures for familiarization and test were nearly identical to those of Exp. 1A. Subjects were quasi-randomly assigned to one of four conditions that determined the order in which participants heard the two grammatical structures and which lexicon each structure was paired with. During the first familiarization phase, participants heard the first language (L1) which utilized the structure of either the adjacent or the non-adjacent language in either the Romance or the Slavic lexicon depending on their experimental condition. The 48 familiarization sentences were presented randomly without replacement three times for an exposure set of 144 sentences during the first familiarization phase, which is half the length of exposure from Exp. 1A. After completing familiarization, subjects immediately began the L1 test phase during which they judged the grammaticality of 24 test strings which were presented randomly using the L1 lexicon (see Appendix). As in Exp. 1A, there were three types of test strings: grammatical familiar, grammatical novel, and ungrammatical. The six grammatical familiar test items (either ABA or AAB depending on the condition) were presented to determine if subjects accepted strings that were heard during familiarization as being grammatical. Subjects' ratings of the six grammatical novel test items (either ABA or AAB depending on condition) were used to determine if they would generalize the structure acquired during familiarization to novel strings that were grammatical but unattested. The 12 ungrammatical strings either did not conform to the grammatical structure of either language (three AAA and three ABB) or conformed to the grammatical structure of the upcoming language in Part 2 (either ABA or AAB depending on the condition). These test items were used to examine how participants rated strings that did not conform to the grammatical structure seen during exposure. None of these test strings contained reduplication. Unlike Exp. 1A, this experiment utilized a continuous scale labelled with "Definitely did NOT come from the language" at the far left and "DEFINITELY came from the language" at the far right. Participants were instructed to select any point on the scale to rate a sentence and to go with their gut reaction as to whether the string might have been something a native speaker of the language would have said when following the rules of the language's grammar.

After completing Part 1, subjects began the second familiarization in which they were exposed to an additional 144 sentences in the second language (L2) which utilized the grammatical structure and lexicon that were not heard in the L1. Immediately afterwards, subjects completed the L2 test which followed the same procedure as the L1 test. This test contained 24 strings (see Appendix) presented randomly without reduplication in the L2 lexicon, including six grammatical familiar, six grammatical novel, and 12 ungrammatical. The ungrammatical test items included three AAA and three ABB strings that did not conform to either the L1 or L2 grammar, as well as six strings that conformed to the grammar of the L1. These L1-conforming strings were of particular interest as their structure conformed to the sentences heard during L1 familiarization while the lexical items were drawn from L2. This allowed us to examine whether participants ratings were influenced by structures observed earlier in the experiment. Since the L2-conforming strings from the L1 test had not yet been encountered by subjects, we also examined whether ratings differed between the ungrammatical string types within the first and second tests.

Finally, subjects progressed to Part 3 which consisted of a test asking participants to rate whether the sentences they heard came from the language in Part 1. This test followed the same procedure as the previous two tests and was implemented to examine whether subjects retained the L1 structure after exposure to the L2 familiarization and test. Subjects were randomly presented with six grammatical familiar strings from the L1 as well as 12 ungrammatical strings (three AAA, three ABB, six L2-conforming). These test strings were followed by the six reduplicated strings (while no reduplication was encountered during L1 familiarization, these strings nonetheless conformed to the L1 grammar). We presented these reduplicated strings to examine whether subjects would rate reduplication similarly to the other ungrammatical test items from the final test (and thus consider them ungrammatical) or to the grammatical familiar strings (similar to the grammatical novel strings from the previous two tests).

Results

We conducted our analyses using R version 3.6.3 (R Core Team, 2016) with the lmerTest package (Kuznetsova, Brockhoff, & Christiansen, 2017) and used orthogonal contrast codes for all categorical variables. For all participants during Exp. 1B ($N = 120$), we conducted several linear mixed-effects models to compare ratings at the three different tests and their interactions between the grammaticality of the test strings, subjects' language history, the grammatical structure heard during familiarization, and the lexicon heard during familiarization. Before conducting our main analyses, we modified our data to remove outliers by eliminating ratings that were more than two standard deviations below mean for the grammatical strings heard during the experiment (familiar and novel) and ratings that were more than two standard deviations above the mean for the ungrammatical test strings (AAA, ABB, opposite structure, and reduplicated). The removal of outliers appeared unidirectional only due to responses being near ceiling level in one direction (upper for grammatical strings and lower for ungrammatical strings).

Learning by Language History

Our primary comparison explored the effect of subjects' language history on ratings throughout the experiment. The first model in this comparison examined the interaction between grammaticality (grammatical vs. ungrammatical test strings), test order (L1 test vs. L2 test), and learners' language history (monolingual vs. bilingual) as a within-subjects fixed effect. For random effects, we added a by-participant random intercept and by-participant random slope for grammaticality and a by-item random intercept. There was a main effect of grammaticality on subjects' ratings ($\beta = 36.61$, $t(79.56) = 5.02$, $p < .001$) such that all participants rated grammatical strings significantly higher than ungrammatical strings in both the L1 and L2. There was also a main effect of test order on rating ($\beta = 5.46$, $t(62.07) = 2.01$, $p = .049$) indicating that there was a general decrease in confidence for ratings from the L1 to the L2. There was no effect of language history (β = -.55, t(118.90) = -.25, p = .81) nor any significant interactions between the fixed-effects. The mean ratings for the different test strings at each of the three testing phases by language history are reported in Table **3-3.**

Table **3-3**: Means and SDs for Ratings during Parts 1, 2, and 3 by Test String and Language History

	Monolingual						Bilingual					
	Part 1 (L1)		Part 2 (L2)		Part 3 (L1)		Part 1 (L1)		Part 2 (L2)		Part 3 (L1)	
Grammatical	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Familiar	76.91	20.04	73.01	20.64	74.32	20.34	78.63	20.70	72.75	20.77	72.71	20.63
Novel	76.18	21.47	72.90	19.68	-	-	77.47	20.99	74.62	20.42	-	-
Ungrammatical												
AAA	68.39	27.06	58.81	28.66	63.34	27.08	67.59	26.42	60.49	27.14	66.21	24.87
ABB	70.86	27.62	65.30	24.10	67.33	25.13	68.77	26.93	65.07	23.42	64.20	26.50
Opposite	61.79	29.18	53.16	28.04	60.02	27.22	61.18	29.54	57.30	27.72	60.16	27.38
Reduplicated	-	-	-	-	38.09	32.70	-	-	-	-	40.92	29.68

Using the same linear mixed-effects model, we modified the variable for grammaticality using orthogonal contrast codes to determine if there were significant differences between ratings of the five different types of test strings seen during the L1 and L2 tests (see Figure **3-1** for visualization of mean ratings for each type of test string between monolingual and bilingual learners). There was not a significant difference between participants' ratings of grammatical familiar and grammatical novel strings (β = .32, t(62.29) = .10, p = .92). We also drew comparisons between the ungrammatical strings, specifically examining whether participants rated the strings that were unlike either language (AAA and ABB) differently from the ones that conformed to the opposite language depending on whether the structure of the opposite string was novel (as when rating the L1 strings) or had been seen before (as when rating the L2 strings). There was not a significant interaction between the ratings of these strings and the order of test (β = .30, t(61.99) = .03, p = .98).

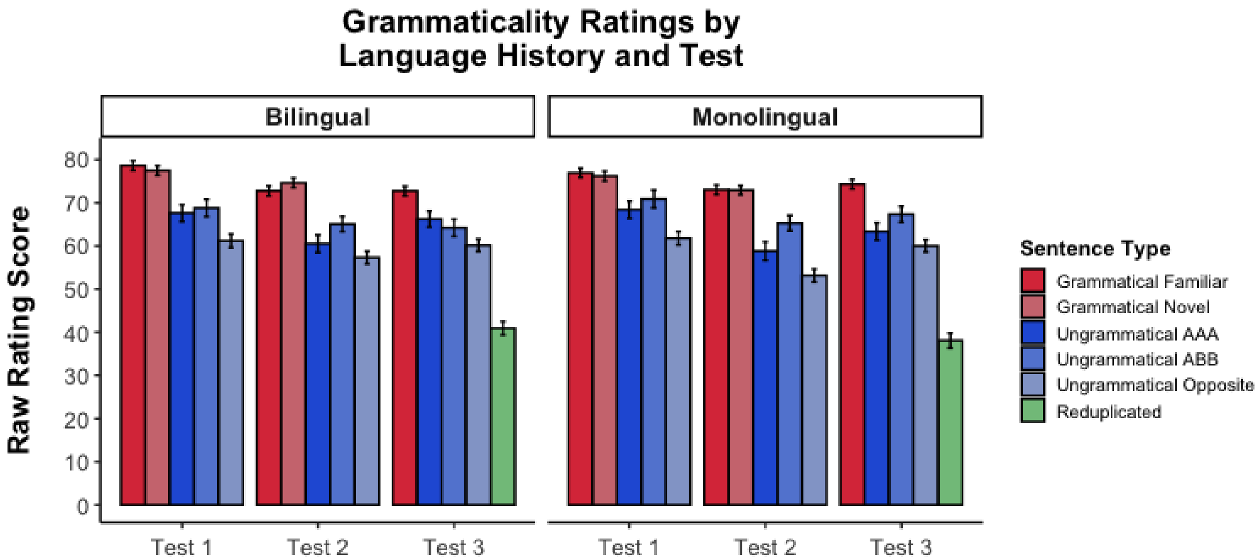

Figure **3-1**: Raw rating scores for test strings from Parts 1, 2, and 3 by language history

The second model in this comparison considered ratings at all three tests. Using orthogonal contrast codes, we focused only on the test strings that were shared across every test, thus removing grammatical novel and reduplicated strings from the model. We looked at the interaction between grammaticality (grammatical vs. ungrammatical test strings), test order (Part 1 vs. Part 2 vs. Part 3 test), and learners' language history (monolingual vs. bilingual) as a within-subjects fixed effect. For random effects, we added a by-participant random intercept and by-participant random slope for grammaticality and a by-item random intercept. We contrasted coded the variable for test order to look at tests 1 and 3 and found a main effect of grammaticality (β = 17.40, $t(77.70) = 4.61$, $p < .001$) as well as a main effect of test order (β = 12.44, $t(64.02) = 5.05$, $p < .001$) on subjects' ratings. When examining tests 2 and 3, there was a main effect of grammaticality (β = 17.39, $t(74.98) = 4.15$, $p < .001$) and a main effect of test order (β = 7.18, $t(64.01) = 2.60$, $p = .01$) with a smaller effect size than when comparing tests 1 and 3, suggesting that ratings tend to decline across the experiment.

Learning by Familiarization Structure

Our second comparison looked at the effect of the grammatical structure heard during familiarization on ratings throughout the experiment. We looked at the interaction between grammaticality (grammatical vs. ungrammatical test strings), test order (L1 test vs. L2 test), and the grammatical structure heard during familiarization (adjacent or non-adjacent) as a within-subjects fixed effect. For random effects, we added a by-participant random intercept and by-participant random slope for grammaticality and a by-item random intercept. There was a main effect of grammaticality on subjects' ratings (β = 36.57, *t*(79.88) = 5.01, *p* < .001) and a main effect of test order on rating (β = 5.46, *t*(62.07) = 2.01, *p* = .049). There was no main effect of grammatical structure (β = -.76, *t*(7537.67) = -1.32, *p* = .18) nor a significant three-way interaction between the fixed-effects. The mean ratings for the different test strings at each of the three tests by familiarization structure are reported in Table **3-4.**

Table **3-4**: Means and SDs for Ratings during Parts 1, 2, and 3 by Test String and Familiarization Structure

	Adjacent (AAB)						Non-Adjacent (ABA)					
	Part 1 (L1)		Part 2 (L2)		Part 3 (L1)		Part 1 (L1)		Part 2 (L2)		Part 3 (L1)	
Grammatical	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Familiar	80.03	19.63	73.54	19.32	77.37	20.45	75.45	20.88	72.21	22.01	69.58	19.80
Novel	80.45	20.24	72.76	19.43	-	-	73.12	21.60	74.78	20.66	-	-
Ungrammatical												
AAA	63.90	30.11	50.80	27.43	63.10	28.02	72.09	22.14	68.50	25.48	66.45	23.77
ABB	77.76	23.11	68.66	20.98	71.58	25.73	61.88	28.80	61.70	25.77	59.95	24.67
Opposite	60.54	30.02	48.23	27.72	60.97	28.95	62.44	62.23	57.30	26.39	59.21	25.52
Reduplicated	-	-	-	-	36.94	33.59	-	-	-	-	42.08	28.51

There was a significant interaction between structure (adjacent vs. non-adjacent) and grammaticality (grammatical vs. ungrammatical test strings) (β = 17.36, *t*(6885.55) = 5.37, *p* < .001). While the simple effect of grammaticality for the adjacent structure (β = 49.53, *t*(99.73) = 6.86, *p* < .001) and the non-adjacent structure (β = 28.62, *t*(99.97) = 3.96, *p* < .001) were both significant, the difference in effect size implies that the variation in ratings between grammatical and ungrammatical strings is significantly greater in the adjacent structure. Within the adjacent structure, the ratings of ABB ungrammatical strings were much higher than the combined ratings for other ungrammatical string types, more closely resembling the ratings for the grammatical strings (see Table **3-4**). We observed the interaction between structure and ABB vs. AAA and opposite strings (β = 13.58, t(8162.43) = 3.91, p < .001) and examined the simple effect of ABB vs. AAA and opposite strings within the adjacent structure (β = 22.12, t(70.77) = 4.15, p < .001), finding that ABB strings were rated significantly higher than all other ungrammatical strings. We observed a similar trend within the non-adjacent language regarding the AAA ungrammatical strings (see Table **3-4**). Looking at the interaction between structure and grammatical vs. AAA strings (β = -10.19, t(8162.43) = -3.91, p < .001), we honed in on the simple effect of grammatical vs. AAA strings within the non-adjacent language (β = 3.89, t(84.25) = 1.50, p = .14) and found that grammatical strings and ungrammatical AAA strings were not rated significantly different within the non-adjacent language (see Figure **3-2** for visualization of mean ratings for each type of test string between the adjacent and non-adjacent structure).

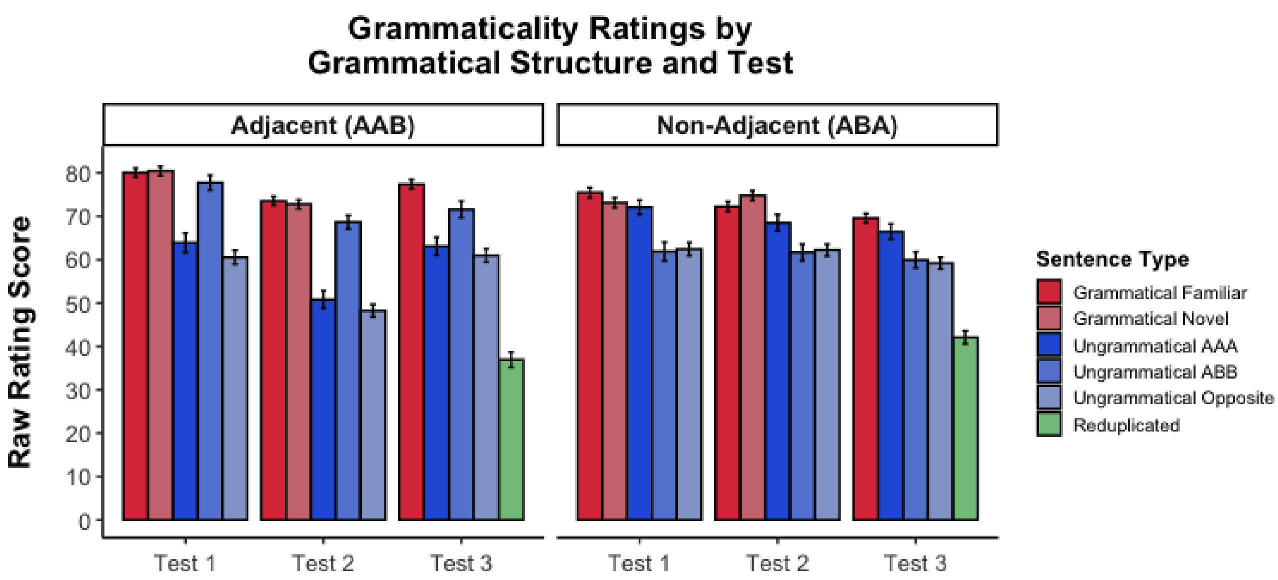

Figure **3-2**: Raw rating scores for test strings from Parts 1, 2, and 3 by grammatical structure

Learning by Familiarization Lexicon

The last comparison examined the effect of the lexicon heard during familiarization on ratings throughout the experiment. We examined the three-way interaction between grammaticality (grammatical vs. ungrammatical test strings), test order (L1 test vs. L2 test) and the lexicon heard during familiarization (Romance or Slavic) as a within-subjects fixed effect. For random effects, we added a by-participant random intercept and by-participant random slope for grammaticality and a by-item random intercept. There was a main effect of grammaticality on subjects' ratings (β = 36.59, t(79.40) = 5.02, p < .001) as well as a main effect of test order on ratings (β = 5.46, t(62.07) = 2.01, p = .049). The mean ratings for the different test strings across all three tests by familiarization lexicon are reported in Table **3-5.**

Table **3-5**: Means and SDs for Ratings during Parts 1, 2, and 3 by Test String and Familiarization Lexicon

	Romance						Slavic					
	Part 1 (L1)		Part 2 (L2)		Part 3 (L1)		Part 1 (L1)		Part 2 (L2)		Part 3 (L1)	
Grammatical	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Familiar	79.57	19.28	69.46	21.17	73.35	21.42	75.90	21.30	76.31	19.64	73.64	19.55
Novel	78.97	20.73	70.48	20.29	-	-	74.71	21.52	76.90	19.34	-	-
Ungrammatical												
AAA	70.63	26.18	56.16	26.74	64.63	26.07	65.35	27.04	63.15	28.63	64.92	26.00
ABB	74.36	26.17	64.02	22.64	66.36	25.35	65.28	27.64	66.35	25.77	65.17	26.36
Opposite	65.31	28.93	51.47	27.21	61.76	26.78	57.67	29.30	58.99	28.18	58.41	27.72
Reduplicated	-	-	-	-	43.00	31.92	-	-	-	-	36.01	30.18

While there was no three-way interaction (β = -2.03, t(403.85) = -.40, p = .69), there was a significant two-way interaction between test order and lexicon (β = 3.81, t(3735.28) = 3.48, p < .001). When examining the simple effect of test order, we found that strings from the L1 test were rated significantly higher than strings from the L2 test within the Romance lexicon (β = 8.93, t(82.41) = 2.83, p = .01) but were not significantly different within the Slavic lexicon (β = 1.49, t(82.45) = .47, p = .64) (see Figure **3-3** for visualization of mean ratings for each type of test string between Romance and Slavic lexicons).

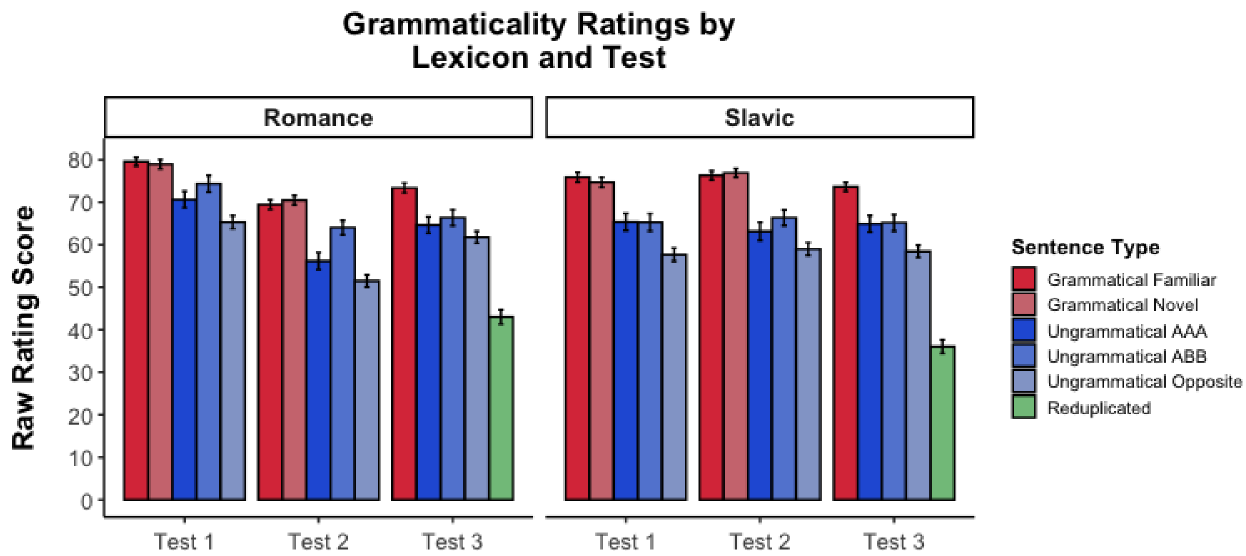

Figure **3-3**: Raw rating scores for test strings from Parts 1, 2, and 3 by lexicon

Reduplicated Strings

During the final test, subjects' ratings of the reduplicated test strings (*M* = 39.51) were much lower than their ratings for ungrammatical strings (*M* = 63.54). To examine whether this difference was significant, we ran a one-way ANOVA on the data from the Part 3 test to assess whether grammaticality (contrast coded as reduplicated vs. ungrammatical strings) impacted participants' ratings. We found that the reduplicated test strings were rated significantly lower than the ratings of ungrammatical test strings during the Part 3 test ($\beta = -18.50$, $F(1, 2848) = 387.91$, $p < .001$).

Chapter 4

General Discussion

In order to explore how monolingual and bilingual learners contend with the statistical learning of grammatical classes when there are multiple languages in the input, we employed a set of two artificial grammar learning experiments based on work by Reeder and colleagues (2013). In Exp. 1A, monolingual English-speaking subjects recruited from the internet were exposed to sentences from one of two artificial languages before completing a grammaticality judgement task. The “Romance” language was comprised of pseudowords based on Spanish and the “Slavic” language was comprised of pseudowords based on Serbian. Both languages contained the underlying grammatical structure (Q)AXB(R), which could only be identified by distributional information embedded in the input. We replicated and expanded on the results of Reeder et al. (2013), finding that online subjects who were exposed to either language rated grammatical sentences significantly higher than ungrammatical sentences during test. They were also able to generalize the grammatical structure to novel stimuli even when the languages contained phonetic and phonotactic properties that were not based on English, the native language of our participants. Notably, these results did not seem to be significantly impacted by the online testing environment as we were still able to replicate the findings from Reeder et al. (2013).

In Exp. 1B, we presented monolingual and bilingual participants recruited online with a task that presented two artificial languages sequentially. One language used the grammatical structure (Q)AAB(R) that placed instances of the A-category adjacent to one another, and the other language used the form (Q)ABA(R) in which the A-categories were non-adjacent. Language structure and use of either the Slavic or Romance lexicons was counterbalanced across participants. After familiarization with the L1, participants completed a grammaticality judgement aimed at identifying whether subjects could retain the structure of the language they just heard. Much like Exp. 1A, monolingual and bilingual learners were able to acquire the grammatical structure of a single language, rating grammatical test stimuli significantly higher than ungrammatical test stimuli. Subjects were then exposed to the L2 followed by another grammaticality judgement task. This task assessed both learning in the L2 and whether subjects would be more accepting of ungrammatical stimuli that conformed to the structure of the L1. We found that subjects were able to acquire the structure of the L2 after being exposed to and tested on the L1. While subjects' ratings were lower overall during the L2 test, they still rated grammatical stimuli significantly higher than ungrammatical stimuli. However, the difference in ratings between these two sentence types was greater during the L1 test. Finally, the experiment concluded with a third test that examined whether learners retained the structure from the first language. This assessment also introduced reduplicated test strings in which both A-categories contained the same lexical item. While reduplicated strings were not heard at any of the familiarization phases, they technically conformed to the grammatical structure of either language and were presented to observe how participants would rate them relative to other ungrammatical strings. Subjects' responses during Test 3 were similar to those of the L2 test, with significantly higher ratings of grammatical stimuli compared to ungrammatical stimuli and lower ratings relative to the L1 test. The reduplicated strings from Test 3 were rated significantly lower than all other ungrammatical strings, suggesting that the highly salient cue of reduplication (e.g., Marcus et al., 1999; Gomez & Gerken, 1999) was observed by participants as deviating from the structure heard during exposure. There did not appear to be any significant differences between ratings by monolinguals compared to those of bilinguals at any of the test phases. When examining test performance by grammatical structure, we found that ungrammatical ABB strings were rated significantly higher in the adjacent language (AAB) than the non-adjacent language (ABA). Additionally, ratings from the non-adjacent language did not vary significantly between the L1 and L2 tests.

Taken together, the findings from Exp. 1B demonstrate that learners are able to acquire the underlying grammatical structure of two artificial languages when they are presented sequentially. Our subjects were able to discriminate between grammatical and ungrammatical stimuli during test in both languages that they heard during the experiment. Further, they retained the grammatical structure of the first language even after being exposed to and tested on the second language. Our participants exhibited a general decrease in confidence during the grammaticality judgment task for all three test phases, rating sentences from the L1 higher than sentences from the L2 or the final test. This pattern is consistent with previous bilingual SL research that has presented learners with multiple patterns in the input. Many of these studies have reported a primacy effect in which participants show increased learning in the first language relative to the second language. A classic example of this effect is Gebhart et al. (2009), who found that when presenting subjects with an auditory stream of structural information, changing the structure changed mid-stream without any additional cues to this change resulted in subjects acquiring only the first structure in the stream. When this changed was marked by a long pause and explicit instructions (stating that subjects would be exposed to multiple structures), they were able to acquire both structures in the input. The primacy effect that we observed in Exp. 1B was related to a general decrease in ratings across the experiment rather than a lack of learning in the second structure. While our subjects were not explicitly cued to the presence of two languages, they had other cues at their disposal that likely aided in learning both languages during the experiment (e.g., phonetic and phonotactic differences between the Romance and Slavic lexicons; the presentation of the L1 and L2 being separated by the L1 test. These results are more closely related to the findings of Poepsel and Weiss (2016), which examined the role of contextual cues in the conscious awareness of learning. Subjects completed a cross-situational statistical learning paradigm in which they were exposed to both 1:1 and 2:1 word-object mappings and were asked to rate how confidently they knew the name of each object. Some subjects were presented with contextual cues during exposure (i.e., change in speaker voice, explicit instructions) that distinguished the 1:1 mappings from the 2:1 mappings while others received no cues. They found that while the presence of contextual cues did not facilitate learning, participants' confidence ratings were significantly higher when there were cues that corresponded to either mapping. While these findings do not directly explain the decrease in ratings observed in Exp. 1B, they do offer a contrast that should perhaps be explored in subsequent research regarding the role of correlated cues in learning confidence.

Our null findings regarding group differences between monolinguals and bilinguals during Exp. 1B further perpetuates the question of how language history impacts statistical learning. A number of studies in the field have yielded mixed results regarding whether there is increased performance in SL tasks by bilingual infants and children as well as adults. In infant learners, the focus of much of this research has been examining whether being raised in a bilingual environment impacts infants' ability to segment speech (e.g., Antovich & Graf Estes, 2017; Bulgarielli et al., 2017) and discriminate between different sets of phonemes (Burns et al., 2007; Sebastián-Gallés & Bosch, 2009), which are challenges that young bilingual learners presumably face early on in language development. While there is evidence that bilingualism may confer an advantage in these areas of statistical learning for infants, these findings seem to be observed less frequently in adult learners (e.g., Kittleson et al., 2010; Yim & Rudoy, 2013). In addition, there is little precedence regarding how monolingual and bilingual adults contend with the learning of grammatical classes as few studies have looked at this question. The lack of group differences observed during Exp. 1B may be a result of robust learning by both groups of participants due to an abundance of available cues.

The observation that participants rated ungrammatical ABB strings significantly higher during the test of the adjacent language (AAB) compared the non-adjacent language (ABA) was not predicted. However, it may be explained in part by previous findings regarding reduplication as a salient cue in language input (e.g., Marcus et al., 1999; Tunny & Altmann, 1999). Though the ABB strings did not conform to either of the grammars presented during the experiment, these strings do contain two of the same categories adjacent to each other. Since our subjects appeared to acquire the underlying structure of the AAB (which also contains two adjacent categories), it is possible that subjects recognized the adjacent categories in the ABB strings and identified them as belonging to the adjacent language even though they did not occur in the same order as the strings heard during familiarization.

There are several limitations of this study, including the sample of bilingual participants recruited during Exp. 1B. We recruited bilingual speakers who spoke English and one additional language, placing some restrictions to screen out those who are familiar with languages that are phonologically similar to Spanish or Serbian. As a result, our population was very heterogenous and our data was sampled from subjects with a wide variety of language backgrounds. The bilingual experience is not easily quantified, and it is challenging to address how variability amongst bilingual speakers (e.g., AoA, proficiency, language use) may affect learning outcomes (see Luk & Bialystok, 2013). Additionally, our sample size for Exp. 1B was rather small with only 30 participants in each the monolingual and bilingual groups (as compared to a total of 60 participants during Exp. 1A) which could have contributed to a lack of between-group differences during our analyses. Due to the nature of the grammatical structures during Exp. 1B, there were only a small number of sentences for each string type to be heard during test. As a result, subjects heard only a few of each ungrammatical string type during the grammaticality judgement tasks (i.e., each test section only had three AAA and three ABB strings). While our mixed effects models did not have issues with convergence, having more observations per cell for each type of test string would have increased the overall power of our model and ensured that participants ratings were not driven by inadvertent biases within specific test items.

The original question posed by Reeder and colleagues (2013) was whether learners could acquire the underlying grammatical structure of an artificial language without the aid of correlated cues. The current research program explores this question as it applies to bilingual statistical learning, asking whether correlated cues are necessary to acquire grammatical classes from two artificial languages. Exp. 1A replicated the results from the original study and vetted the validity of the Romance and Slavic lexical items, while Exp. 1B showed that learners could acquire the structure of two languages presented sequentially with the aid of several correlated cues. Future research will further investigate the role of correlated cues in the distributional learning by reducing the number of cues that could aid subjects in acquiring grammatical classes. An upcoming study will draw from the design of Gebhart et al. (2009), presenting subjects with two languages sequentially in a single familiarization phase without a break or explicit cue to the change in structure. If subjects are able acquire the structure of both languages, we will continue to strip away correlated cues during exposure and assess learning when there are no phonetic and phonological differences between languages.

References

Antovich, D. M., & Graf Estes, K. (2018). Learning across languages: Bilingual experience supports dual language statistical word segmentation. *Developmental Science*, *21*(2), e12548.

Boersma, Paul & Weenink, David (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.40, retrieved 27 February 2021 from http://www.praat.org/

Braine, M. D. S. (1987). What is learned in acquiring word classes – A step toward an acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 65–87). Hillsdale, NJ: Lawrence Erlbaum Associates.

Braine, M. D. (1966). Learning the positions of words relative to a marker element. *Journal of Experimental Psychology*, *72*(4), 532.

Bulgarelli, F., & Weiss, D. J. (2016). Anchors aweigh: The impact of overlearning on entrenchment effects in statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1621.

Bulgarelli, F., Benitez, V., Saffran, J., Byers-Heinlein, K., & Weiss, D. J. (2017, May). Statistical learning of multiple structures by 8-month-old infants. In *Proceedings of the... Annual Boston University Conference on Language Development. Boston University Conference on Language Development* (Vol. 41, p. 128). NIH Public Access.

Bulgarelli, F., Lebkuecher, A. L., & Weiss, D. J. (2018). Statistical learning and bilingualism. *Language, speech, and hearing services in schools*, *49*(3S), 740-753.

Bulgarelli, F., Bosch, L., & Weiss, D. J. (2019). Multi-pattern visual statistical learning in monolinguals and bilinguals. *Frontiers in psychology*, *10*, 204.

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*, *28*(3), 455-474.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, *87*(2), B69-B77.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Crystal, D. (2011). From the world to the word-and back again. *Plenary lecture given to the CILT Primary Languages Show,'From the Word to the World', Liverpool, 4 March 2011*.

Culicover, P. W., & Culicover, P. W. (1999). *Syntactic nuts: Hard cases, syntactic theory, and language acquisition* (Vol. 1). Oxford University Press on Demand.

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128.

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive science*, *33*(6), 1087-1116.

Gleitman, L. R., & Wanner, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3–48). Cambridge, England: Cambridge University Press.

Gleitman, L. R. (1990). The structural sources of verb meaning. Language Acquisition, 1, 3–55.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109-135.

Grimshaw, J. (1981). Form, function, and the language acquisition device. In C. L. Baker & J. J. McCarthy (Eds.), *The logical problem of language acquisition* (pp. 165–187). Cambridge, MA: MIT Press.

Grosjean, F. (2008). *Studying bilinguals*. Oxford University Press.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior research methods*, *42*(3), 627-633.

Keuleers, E. (2015). vwr: Useful functions for visual word recognition research. R package version 0.3.0.

Kittleson, M. M., Aguilar, J. M., Tokerud, G. L., Plante, E., & Asbjørnsen, A. E. (2010). Implicit language learning: Adults' ability to segment words in Norwegian. *Bilingualism: Language and Cognition*, *13*(4), 513-523.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, *82*(13), 1-26.

Labelle, M. (2005). The acquisition of grammatical categories: The state of the art. *Handbook of categorization in cognitive science*, 433-457.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*(5), 605-621.

Maratsos, M., & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. Nelson (Ed.). Children's language (Vol. 2, pp. 127–189). New York: Gardner Press.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77-80.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental science*, *11*(1), 122-134.

Pinker, S. (1984). Language learnability and language development. Cambridge: Harvard University Press.

Pinker, S., & MacWhinney, B. (1987). The bootstrapping problem in language acquisition. *Mechanisms of language acquisition*, 399-441.

Poepsel, T. J., & Weiss, D. J. (2014). Context influences conscious appraisal of cross situational statistical learning. *Frontiers in psychology*, *5*, 691.

Poepsel, T. J., & Weiss, D. J. (2016). The influence of bilingualism on statistical word learning. *Cognition, 152*, 9-19.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of verbal learning and verbal behavior*, *6*(6), 855-863.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive science*, *22*(4), 425-469.*Cognition, 152,* 9-19.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology, 66(1)*, 30-54.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928.

Sebastián-Gallés, N., & Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants: Is the distributional account all there is to it?. *Developmental science*, *12*(6), 874-887.

Serratrice, L. (2013). Cross-linguistic influence in bilingual development: Determinants and mechanisms. *Linguistic Approaches to Bilingualism*, *3*(1), 3-25.

Team, R. C. (2013). R: A language and environment for statistical computing.

Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2009). Speech segmentation in a simulated bilingual environment: A challenge for statistical learning?. *Language Learning and Development*, *5*(1), 30-49.

Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*(7), 2064-2072.

Yim, D., & Rudoy, J. (2013). Implicit statistical learning and language skills in bilingual children. *Journal of Speech Language and Hearing Research, 56*(1), 310–322.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological science*, *18*(5), 414-420.

Appendix

Test Strings

Exp. 1A: AXB Strings Used During Test

Grammatical Familiar			Grammatical Novel			Ungrammatical					
A_1	X_1	B_3	A_1	X_1	B_2	A_1	X_1	A_2	B_2	X_1	B_1
A_2	X_1	B_2	A_2	X_1	B_1	A_1	X_2	A_3	B_3	X_2	B_1
A_3	X_1	B_1	A_3	X_1	B_3	A_1	X_3	A_3	B_3	X_3	B_1
A_1	X_2	B_2	A_1	X_2	B_1	A_2	X_1	A_3	B_1	X_1	B_2
A_2	X_2	B_1	A_2	X_2	B_3	A_2	X_2	A_1	B_1	X_2	B_2
A_3	X_2	B_3	A_3	X_2	B_2	A_2	X_3	A_1	B_1	X_3	B_2
A_1	X_3	B_1	A_1	X_3	B_3	A_3	X_1	A_1	B_1	X_1	B_3
A_2	X_3	B_3	A_2	X_3	B_2	A_3	X_2	A_2	B_2	X_2	B_3
A_3	X_3	B_2	A_3	X_3	B_1	A_3	X_3	A_2	B_2	X_3	B_3

Exp. 1B: AXB Strings Used During Part 1 and Part 2 Test

Grammatical Familiar			Grammatical Novel			Ungrammatical					
Nonadjacent Language											
A_1	B_1	A_2	A_2	B_1	A_3	A_2	A_1	A_3	A_1	A_3	B_1
A_2	B_1	A_1	A_3	B_1	A_1	A_3	A_2	A_1	A_3	A_2	B_1
A_1	B_2	A_2	A_2	B_2	A_1	A_1	A_3	A_2	A_2	A_3	B_2
A_1	B_2	A_3	A_3	B_2	A_2	A_2	B_1	B_2	A_3	A_1	B_2
A_2	B_3	A_3	A_1	B_3	A_3	A_3	B_2	B_3	A_1	A_2	B_3
A_3	B_3	A_1	A_3	B_3	A_2	A_1	B_3	B_1	A_2	A_1	B_3
Adjacent Language											
A_1	A_3	B_1	A_2	A_3	B_1	A_2	A_1	A_3	A_1	B_1	A_2
A_3	A_2	B_1	A_3	A_1	B_1	A_3	A_2	A_1	A_2	B_1	A_1
A_2	A_3	B_2	A_2	A_1	B_2	A_1	A_3	A_2	A_1	B_2	A_2
A_3	A_1	B_2	A_3	A_2	B_2	A_1	B_1	B_3	A_1	B_2	A_3
A_1	A_2	B_3	A_1	A_3	B_3	A_2	B_2	B_1	A_2	B_3	A_3
A_2	A_1	B_3	A_3	A_2	B_3	A_3	B_3	B_2	A_3	B_3	A_1

Exp. 1B: AXB Strings Used During Part 3 Test

Grammatical Familiar			Reduplicated			Ungrammatical					
Nonadjacent Language											
A_1	B_1	A_3	A_1	B_1	A_1	A_1	A_2	A_3	A_1	A_2	B_1
A_3	B_1	A_2	A_2	B_1	A_2	A_2	A_3	A_1	A_2	A_1	B_1
A_2	B_2	A_3	A_2	B_2	A_2	A_3	A_1	A_2	A_1	A_2	B_2
A_3	B_2	A_1	A_3	B_2	A_3	A_1	B_2	B_3	A_1	A_3	B_2
A_1	B_3	A_2	A_1	B_3	A_1	A_2	B_3	B_1	A_2	A_3	B_3
A_2	B_3	A_1	A_3	B_3	A_3	A_3	B_1	B_2	A_3	A_1	B_3
Adjacent Language											
A_1	A_2	B_1	A_1	A_1	B_1	A_1	A_2	A_3	A_1	B_1	A_3
A_2	A_1	B_1	A_2	A_2	B_1	A_2	A_3	A_1	A_3	B_1	A_2
A_1	A_2	B_2	A_2	A_2	B_2	A_3	A_1	A_2	A_2	B_2	A_3
A_1	A_3	B_2	A_3	A_3	B_2	A_1	B_2	B_3	A_3	B_2	A_1
A_2	A_3	B_3	A_1	A_1	B_3	A_2	B_3	B_1	A_1	B_3	A_2
A_3	A_1	B_3	A_3	A_3	B_3	A_3	B_1	B_2	A_2	B_3	A_1