Using Eye-Tracking to Examine Grammatical Predictability in Spanish-English Bilinguals and Spanish Language Learners

by

Guadalupe de los Santos

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in the University of Michigan
2017

Doctoral Committee:

Professor Julie E. Boland, Chair
Assistant Professor Jonathan R. Brennan
Professor Nick C. Ellis
Professor Richard L. Lewis

Guadalupe de los Santos

gdelossa@umich.edu

ORCID iD: 0000-0002-8195-6682

ACKNOWLEDGEMENTS

This dissertation was only possible because of all the people who advised, helped, and supported me. First, I would like to thank my doctoral committee. My advisor, Julie Boland, whose input and advice were invaluable throughout the course of the project, from the inception to the dissertation manuscript. Rick Lewis, whose advice with mixed-effects models was vital, as well as for allowing me the use of his lab and eye-tracker. Jon Brennan also provided valuable statistics advice, especially close to deadlines; in addition, he provided a wealth of linguistics information and background. Nick Ellis provided unique insight into second language learning, which allowed me to extend the project into a new population. I would also like to thank the entire committee as a whole for being willing to meet with me throughout the last year; your feedback and advice helped guide me in this project.

Other people also provided valuable input into the project. Mike Shvartsman was involved in the initial conception of the project, and helped with the eye-tracking programs; Lorenzo García-Amaya provided useful insight into language proficiency testing and allowed me to use a Spanish proficiency test he had developed. The psycholinguistics lab provided wonderful feedback throughout the years. I thank you all.

I would also like to thank Dave Meyer, whose support and encouragement convinced me I could finish the dissertation. I could not have done this without your positivity and belief in me.

My friends encouraged me from start to finish, especially Preeti Samudra and Jenny Weiss; your feedback and advice helped on every step of the project. Finally, last but not least, I would like to thank my family. Your unconditional love and support made this possible—I cannot thank you enough.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Bilingualism is prevalent, with over half of the population of the world being bilingual. While bilinguals have traditionally been viewed as having two separate languages, modern views of language suggest that languages are not completely separate in the mind. This is especially evident in cases of intrasentential code-switching, when a speaker switches languages mid-sentence. Such points are of interest because they represent cases when the languages are activated simultaneously.

This dissertation expands our understanding of multi-language representation by investigating whether some grammatical representations generated by Spanish-English bilinguals and Spanish L2 language learners during reading are specific to the input language used to create the representation, or whether those representations are language-independent. Using eye-tracking, we measured reading times on nouns in grammatical (determiner-noun) and ungrammatical (adverb-noun) contexts, in both same language and mixed language pairs, as participants performed a two-string lexical decision task.

Experiment 1 found that bilinguals read nouns faster following determiners than adverbs. Crucially, this grammatical predictability effect did not interact with the same/mixed language variable. This suggests that grammatical predictability in this context is language-independent, not affected by language nor the presence of a language switch. Experiment 2 found a similar pattern for Spanish language learners, though it was not significant. When the data for Experiment 1 and Experiment 2 were combined, there was a main effect of grammaticality that did not interact with language congruency, suggesting that language-independent predictions influenced reading times for both bilinguals and language learners.

Experiment 3 took into account categorical ambiguity, i.e., that the same word can belong to more than one grammatical class. We computed two conditional probabilities over abstract grammatical categories to represent grammaticality in a more fine-grained way, allowing syntactic category ambiguity. Participants read the second word faster as its probability given the category of the first word increased. This grammatical predictability effect was language-independent, in that it was not modulated by a language switch.

Overall, this dissertation provides an in-depth investigation into multi-language representation and grammatical predictability in Spanish/English bilinguals, focusing on syntactic sequences that have the same word order in the two languages. Our results most strongly support the shared syntax view of bilingual language representation, having found language-independent grammatical predictability across experiments.

Chapter 1

Previous Research and Overview of Dissertation

Introduction

Bilinguals produce and understand sentences in both their languages, but how independent are the language systems? Converging evidence supports the idea that linguistic knowledge is shared between languages. For example, several syntactic priming studies have found that, after encountering a grammatical form (e.g., a passive) in one language, bilinguals were more likely to produce sentences with the same form in their other language (Hartsuiker et al., 2004; Loebell & Bock, 2003). Similar cross-language effects have also been found in comprehension, using a variety of tasks (Meijer & Fox Tree, 2003; Shin & Christianson, 2009; Weber & Indefrey, 2009). This prior research suggests that grammatical representations are shared across languages, but has focused on language changes (i.e. code-switches) across sentences.

Our concern is with incremental word-by-word generation and immediate subsequent use of grammatical representations: our working hypothesis is that bilingual readers' representations are shared enough across language to affect processing across within-sentence code-switch. At some level, this hypothesis is obviously true, given that bilinguals produce and understand mid-sentence code-switches. However, there has been little research devoted to the online, incremental processing of such code-switches during comprehension. Specifically, the prediction was that facilitation of predictable words and grammatical completions (Rayner, 1998) will persist immediately across a switch. Eye-tracking was used to examine word-by-word processing times of grammatical and ungrammatical word pairs at points of language switches within possible grammatical constituents.

In the remainder of this chapter, we will first look at previous research looking at the independence of grammatical representations across languages, focusing on bilinguals. We will then cover some theories of bilingual language representation, before moving onto the methodology and experiments that make up this dissertation. It is important to note that we use the term bilingual to mean a person who learned both their primary languages before puberty (this is not always consistent across the literature). We were also interested in language representation among language learners (people who learned their second language past puberty), but will not look at previous research and theories concerning language learners until Chapter 3.

Code-switching

One feature often observed among bilinguals is the fact that they often *code-switch*, i.e., switch from speaking one language to another, when speaking to other bilinguals in an appropriate social setting. Code-switching occurs for many types of reasons; such as cognitive, when a person has difficulty finding the right word, or social, to foster a sense of community, or it might be encountered for other reasons such as economic, when companies use it in adverts (Basnight-Brown & Altarriba, 2007). Code-switching has been observed at all ages, from children to adults. Though at first thought to be the result of improper language differentiation, studies have shown that bilinguals (even young toddlers) are sensitive to their audience and match their language use to that of their interlocutor (Genesee et al., 1995). Such research has shown that code-switching is not due to a lack of linguistic ability, but rather it is a function of a bilingual's language systems.

Code-switching can occur either within or between sentences. Though code-switches across sentences are generally accepted, not all intrasentential (within sentence) code-switches are universally accepted. This implies that some linguistic constraints must be governing code-switches (Gringràs, 1974; Pfaff, 1979). The nature of the constraints is under debate; though the goal is a universal model of code-switching, models are based on available data and case studies, each looking at particular set of languages. What seems to be a constraint in one language-pair may not hold when looking at a different language pair.

The model of code-switching we are most concerned with was proposed by Sankoff and Poplack, and argued that code-switches can occur only when both languages share a common word order (equivalency) and between free morphemes (1981). These constraints came from research on naturally occurring Spanish-English code-switches; switches occur at points where languages share a common word order, preserving grammaticality in both languages (Poplack, 1980). The main reason we are focusing on this model is due to the fact that we are using Spanish-English bilinguals, and so our participants will follow the patterns that led to the formation of these principles. More discussion on the model and its limitations will be covered in the introduction to Experiment 1.

In regards to our study, code-switching represents a point during natural language use when a bilingual has both languages activated. We are interested in investigating how a bilingual's languages are processed, and code-switching provides us with natural data of languages being activated concurrently.

Syntactic Priming

Syntactic priming is a powerful tool for investigating syntactic representations. Listeners are more likely to produce the syntactic structure that they have just encountered. In the classical picture-description task, participants hear a sentence, repeat it, and then describe an image shown to them. For example, Bock (1986) found that if a listener heard '*the man gave her the ball*' and was shown an image that could be described as either '*the woman showed the girl the book'* or '*the woman showed the book to the girl*', the listener would be more likely to use the same double object [DO] dative structure that they had just heard than the prepositional dative [PO] structure (i.e., they were more likely to repeat *'the woman showed the girl the book'*). If, instead, they heard '*the man gave the ball to her*', they would be more likely to use the PO structure than the DO structure.

Cross-linguistic syntactic priming. Syntactic priming has also been found to occur across languages. Loebell and Bock (2003) first examined the effect of cross-linguistic syntactic priming using German-English bilinguals and a picture description task. They focused on the dative structure, which is similar in English and German, having both a DO structure and a PO structure, each having the same word orders in both languages for each respective alternation. They found similar levels of syntactic priming both within and across languages with the dative structure alternations. However, when using passive transitive primes in German, looking at what English sentences were produced, they found no evidence of cross-linguistic syntactic priming. Since active transitives share the same word order in German and English, but passive transitives do not, it supports the idea syntactic structures can be shared, but only when the word order is the same (Loebell & Bock, 2003). Previous evidence weakens the claim, however, due to the fact that passive transitive priming is not always found when using German passive primes, even when looking at monolinguals (Bock, 1986).

Other cross-linguistic syntactic priming studies with a picture description task also showed mixed results. Same word order was found to be necessary when looking at relative causes in German, Dutch, and English; priming was found when the word order was the same, in German and Dutch relative clauses—but not when the word order was different—English and Dutch relative clauses (Bernolet et al., 2007). In contrast, shared word order was not found to be necessary when looking at transitives in English and Chinese, where priming was found though the word order was different (Chen et al., 2013). Cross-linguistic syntactic priming unaffected by word order was also found in a variety of other tasks: in a sentence recall task (Shin & Christianson, 2009), a sentence completion task (Desmet & Declercq, 2006), and in neural activity repetition suppression (Weber & Indefrey, 2009).

Causes of syntactic priming. Syntactic priming is a powerful tool because it allows us to investigate syntactic representations, and more relevantly to our interests, it allows us to probe at how syntactic representations may be shared across languages. Two possible explanations are usually given for why syntactic priming occurs. The implicit learning account states that when we use a structure, we are unconsciously reinforcing it, which makes it more likely that we use this structure in the future (Bock, 1990; Chang et al., 2006). This account is appealing because it helps explain long-term effects that have been found using syntactic priming. Syntactic priming effects have been found with numerous trials between the prime and target (Bock & Griffin, 2000) and there has even been some evidence that suggests that priming effects can last one week past the original exposure to the priming structure (Kaschak et al., 2011). One drawback of the implicit learning account however, is that it fails to provide an explanation as to why we may see cross-linguistic structural priming when the word order is different between the prime and target of a sentence.

Another possibility is that syntactic priming is lingering activation of a previously encountered structure (Pickering & Branigan, 1998). In this account, a structure is activated when encountered during the prime; it is still in the short-term memory when it comes time to produce the target. For example, in the picture description task example described at the beginning of the section, the DO alternation would be activated in the participant when they hear and repeat the prime sentence. They now have the DO form activated prior to forming the new sentence describing the target image, and so are more likely to use the DO form when producing it than they would have been otherwise.

Though unable to explain long-term effects found in syntactic priming, the lingering activation account has several advantages. First, it suggests that syntactic priming should not be limited to language production. If syntactic priming is similar to other implicit learning activities, the learning occurs in production; we should not see it in language comprehension. This is not the case, as syntactic priming has been found in a variety of tasks not involving language production, such as sentence recall (Potter &

Lombardi, 1998), attachment preference (Branigan et al., 2005), resolving sentence ambiguities (Thothathiri & Snedeker, 2008), ERPs during the reading (Ledoux et al., 2007), and grammaticality assessments in reading (Luka & Barsalou, 2005).

Second, it can explain why there is syntactic priming when the word order is different, as we can find across languages. By the lingering activation account, rather than activating the exact word order, what is activated is a more abstract functional form, which remains activated across languages (Hartsuiker el al, 2004). For example, in the priming of a passive transitive structure, the 'passive' is what is being primed, the idea that the direct object is the subject of a sentence rather than the actor. These findings have led to a change in bilingual language theories, as we will see in the following section.

Theories of Bilingual Syntactic Representation and Processing

Theories of bilingual syntactic representation need to explain why a sentence prime in one language can lead to increased production of the same structure even across languages. They also need to provide an explanation as to how and why code-switching might occur, as it has been found among many groups of bilinguals. One question often focused upon is the degree of overlap between the languages. Different models have different predictions based on how the languages are thought to be represented.

Bilingual production model. One influential model that postulates little overlap in grammatical encoding is de Bot's bilingual adaptation of Levelt's speaking model (1992). In this bilingualism production model, each language has its own conceptualizer, where knowledge is first translated into a preverbal message to be changed into speech—separate systems are noted to be necessary due to the fact that different languages encode some concepts differently. For example, English has only two terms to refer to the distance of items in relation to the speaker, near the speaker (*'here'*) or further away (*'there'*). In comparison, Spanish has three, one referring to items near the speaker ('*aquí*'), one referring to items far away (’*allí*’), and one referring to items to items a medium distance away (’*ahí*’). The preverbal message is then transformed into words at a language-specific formulator, receiving language-specific grammatical encoding, and getting words from a shared lexicon. Figure 1 shows a visual representation of de Bot’s model.

Grammatical encoding occurs in the formulator, for which there exists one per language—having separate formulators reduces the need for coordination between languages. While advantageous in that it removes the issue of coordination, this account is vague as to how code-switching might occur. The formulators are connected through a shared lexicon, since the ‘meaning’ part of a lemma has no language. Though the model proposes separate formulators, it does not rule out the possibility that the formulators may interact in some manner, though it is unclear how it would do so. It allows for activation in one formulator leading to weaker activation in the other. This would allow for cross-linguistic syntactic priming; however, it predicts that within-language syntactic priming effects should be stronger than cross-linguistic syntactic priming effects. One problem for this theory is that numerous studies have found that cross-linguistic syntactic priming is just as strong as within-language syntactic priming (Harsuiker et al., 2004; Kantola & van Gompel; Schoonbaert et al., 2007). The current bilingual production model cannot adequately account for all cross-linguistic priming phenomena, and though code-switching can theoretically occur, it does not provide an adequate explanation as to why it does.

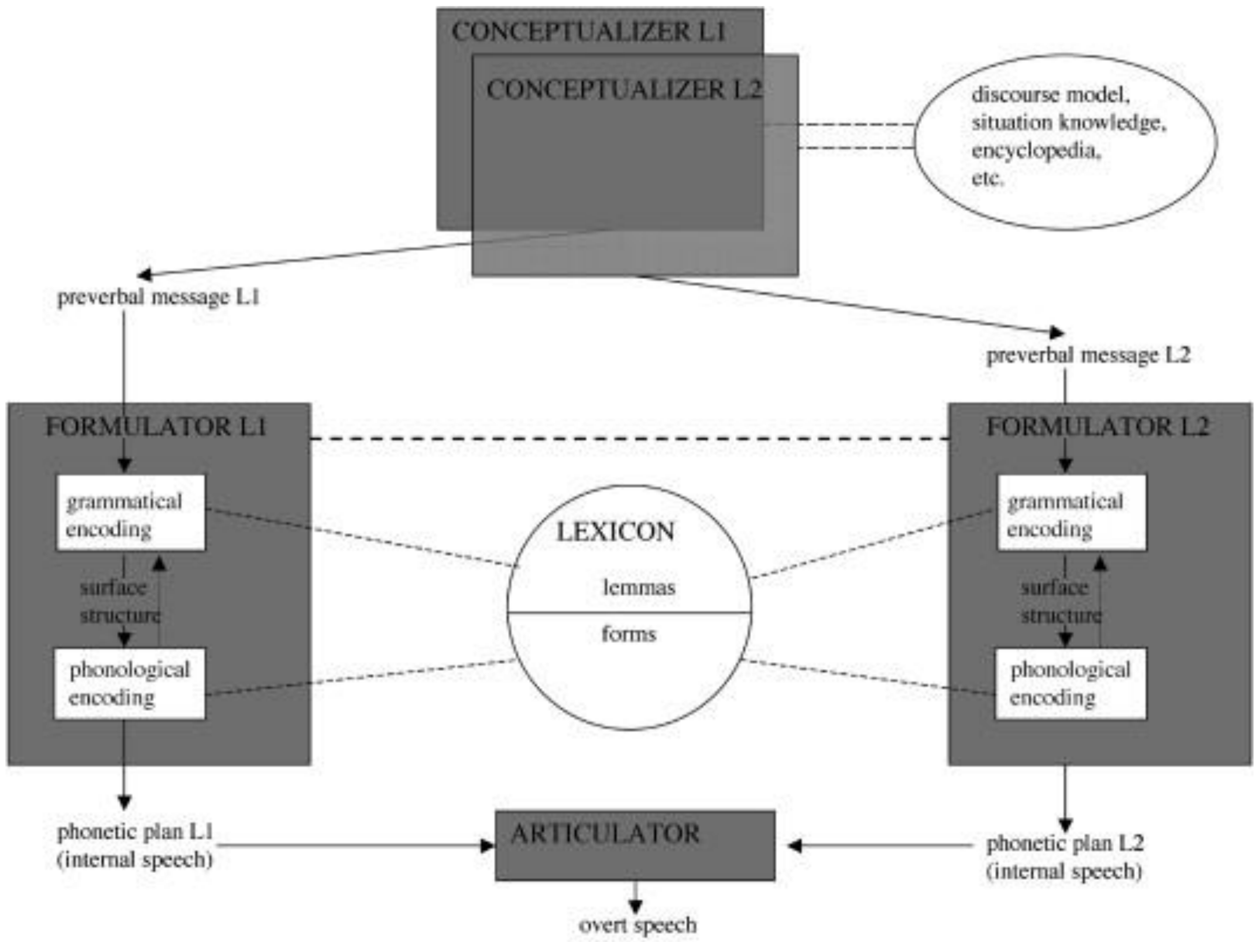


Figure 1: Representation of de Bot's bilingual production model. Lexicon is a shared space, but otherwise separate systems for each language [reprint from Hartsuiker & Pickering, 2008].

Shared syntax account. An alternate theory of bilingual language representation proposed is the shared syntax account (Hartsuiker et al., 2004). Similar to de Bot's bilingual production model, this theory posits a shared lexicon. Unlike the bilingual production model, the shared syntax account focuses on how syntax interfaces with the lexicon. Furthermore, the model accounts for both production and comprehension, as connections are bidirectional. Lexical entries are divided into three levels of information: the conceptual, lemma, and a word form level. Syntactic information is connected at the lemma level. Figure 2 shows a small portion of how a Spanish-English bilingual might represent transitive verbs in the shared syntax model.

Conceptual information includes the meaning of the concept, including the necessary agents/objects. In the figure, the conceptual level is indicated as the highest two nodes [hit and chase], with each concept taking two arguments (an agent and an object). Each conceptual node is connected to the lemma nodes of the words as represented in a bilingual's languages. The lemma node has several important connections; first to a node indicating its language (represented by the British or Spanish flags); second, to its syntactic category (verb), and finally, to combinatorial nodes—active and passive.

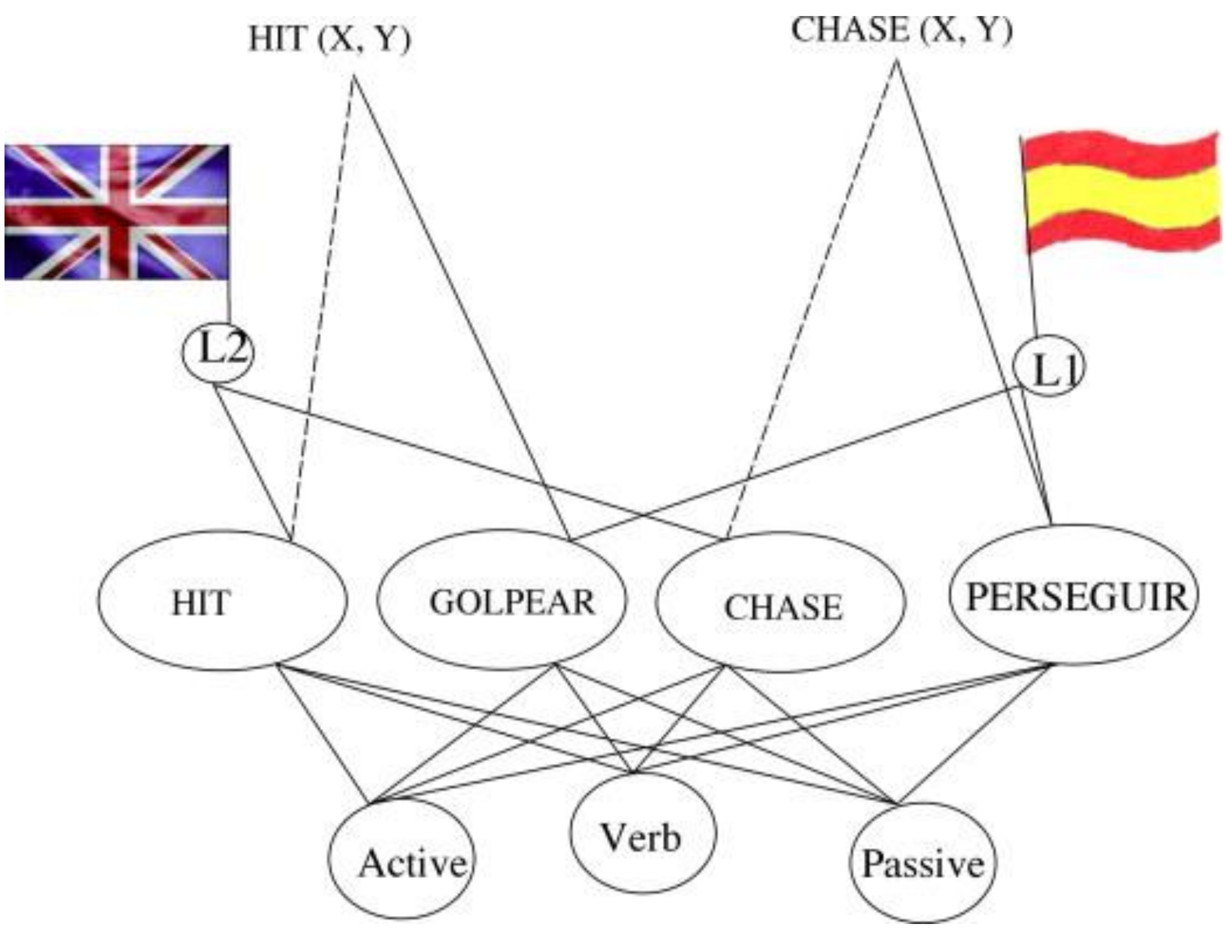


Figure 2: Representation of shared syntax account. Figure shows a partial representation of the verbs 'hit/golpear' and 'chase/perseguir' at the lemma level in a Spanish-English bilingual [reprint from Hartsuiker & Pickering, 2008].

The activation of a lemma in one language, e.g., 'hit', leads to the activation of its translational equivalent, 'golpear'. Since 'hit' and 'golpear' are the same grammatical category—'verb', and are linked to the same combinatorial nodes due to the fact that they are both transitive verbs, using 'hit' in a sentence will also lead to further activation of 'golpear' through the category and combinatorial nodes.

The shared syntax account has the advantage of being able to account for such phenomena such as cross-linguistic syntactic priming, since activation spread through the combinatorial nodes to the lemma nodes before reaching the language nodes. Similarly, code-switching is also a natural consequence. The above figure only describes the verb, but a transitive verb requires a direct object (the 'Y' in the conceptual node). Syntax is spread at the lemma level, bypassing language, and so the verb and direct object need not be in the same language.

Word-by-word comprehension. Though both the bilingual production model and the shared syntax account provide a framework describing how two languages might interact, both models lack detail on what is going on during word-by-word language processing. The bilingual production model focuses on how a bilingual produces sentences in their various languages rather than how comprehension might occur. Given that languages are represented mostly separately, with a shared lexicon, a possible extension into comprehension is that languages are decoded separately; specific features might cue the listener/reader as to which language they are interpreting, and so they interpret according to that language 'mode'. Shifting control between different grammatical encoders/decoders could account for different language modes. Language mode has been a proposed mechanism to account for how bilinguals act in different situations (Grosjean, 2001). However, while this may explain overall patterns in the production and interpretation of utterances by bilinguals, it is not clear how this works on a word-by-word level.

Similarly, the shared syntax account also fails to account for what occurs on a word-by-word level. The model focuses on the sentence-level, and is vague as to what information is encoded in the combinatorial nodes and lemma nodes. The model is clear in that lemmas from both languages are activated simultaneously, meaning the syntactic information from both languages is activated, as that is encoded in the lemma nodes. Syntactic and semantic information is also present in the category and combinatorial nodes; however, the step-by-step process of how language interacts with these nodes is vague, especially since language is a step removed and only connected to the lemma nodes.

Eye Movements and Eye-tracking

Eye-tracking has long been used in research as a way to determine with great temporal and spatial precision what a person is visually processing at a given time. Posner (1980) discussed how one's eye-movements are linked to attentional processes, and therefore can be used as an effective way of examining what a person is thinking or processing at the time. Eye-tracking methods have been used to investigate language, and particularly written language processing, since the 1970s (Rayner, 1998). Studies have looked at different phenomena using eye-tracking; of most relevance to our study, language processing when reading. In the following sections, we discuss the eye-tracking research on basic reading patterns, spillover effects, word length, word frequency, and syntactic predictability. Across all these areas, longer fixation times are correlated with higher cognitive load (Just & Carpenter, 1980).

Reading patterns. As we read, our eyes are not constantly moving. Instead, our eyes jump from fixation to fixation. Average fixations last approximately 200-250 milliseconds in proficient English readers, and these fixations are followed by a saccade (jump to the next fixation) to 7-9 letter spaces away on average (Rayner, 1984). Given these patterns, not every word is fixated upon. In particular, short words are skipped, and research has found a negative association between word length and the probability that a person would fixate on the word, with words less than four letters long being fixated on only 25% of the time (Rayner & McConkie, 1976). This results in function words, that tend to be very short, being skipped fairly often, only being fixated on 35% of the time (Carpenter & Just, 1983).

Most reading in English (and Spanish) occurs spatially from left to right. However, 10-15% of the time, readers make regressions to earlier positions in the text (Kennedy & Murray, 1987). Regressions can occur for various reasons including comprehension difficulties and needing to fixate on parts of words that were skipped in a long saccade. In the current experiment, we will be presenting two (word or nonword) letter strings on the same line, widely spaced, to encourage fixations on both words. The critical word is the second letter string, and in the critical conditions it is always preceded by a real word.

Word length. Another factor that is strongly related to how long it takes to read a word is how long the word is. Unsurprisingly, it has been consistently found that longer words are read slower than shorter words. Some research has suggested that the relationship between word length and reading time may be linear, where reading time increases per character added at a steady rate (Vitu et al., 1990). Word length effects are quite strong in reading, and therefore, in the present dissertation study design, we equated average Word 2 length across conditions, and included individual Word 2 length as a fixed factor in the analyses.

Word frequency. Research has shown that words that are more commonly occurring in a language are read faster than words that are less frequent. Unlike word length, this effect is not linear, but rather logarithmic (Howes & Solomon, 1951). Word frequency effects, much like word length effects, strongly affect reading times. In this study, average Word 2 frequency was equated across conditions, and the log-transformed individual word frequencies were included as a fixed factor in the analyses.

Spillover effects. Spillover effects can be divided into frequency effects and length effects. For the former, words are read faster following a more frequent word compared to a less frequent word when all other factors are controlled for (Kennedy, 1998). For example, you would expect the word ‘house’ to be read faster following the word ‘the’ (frequency: 51,764/million) rather than the word ‘any’ (frequency: 780/million). This is an effect that is independent of syntax, and has exclusively to do with how commonly used the word read before the current word is in the relevant language (Inhoff & Rayner, 1986). For spillover length, much like with spillover frequency, words are read faster following shorter words as compared to longer words when controlling for all other factors (Kennedy et al., 2002).

In the present studies, we were unable to control for the length and frequency of Word 1, making spillover effects a potential concern. Therefore, we will include both log-transformed Word 1 frequency and Word 1 length as fixed factors to ensure spillover effects do not impact the results of interest.

Predictability of words. As has been discussed earlier, words in multi-word text are not processed in isolation, and reading times can be impacted by words occurring previously in the text. A critical feature of central importance in the present dissertation is how predictable a word is based on the previous word read. Research on how the predictability of words influences reading time is discussed in more detail in the following section.

Predictability

Expected words are more quickly processed than unexpected words during language comprehension (Ashby et al., 2005; Rayner et al., 2004). Expectation can be seen in a variety of contexts, including phonetics (Allopenna et al., 1998), semantic content (Becker, 1980), and syntactic context (Goodman et al., 1981). The present study focuses on predictability effects in syntax. Below, we focus on several lines of research looking how syntax affects predictability.

Misleading sentences. One area in which predictability effects are seen is in the resolution of misleading sentences. Misleading sentences are sentences that contain an initial ambiguity that readers are predisposed to interpret in one way, but which is later shown to have a different structure. One type of misleading sentences are garden path sentences (so named because the reader is mislead, or sent down a garden path). For example, when encountering a sentence such as *'Since Jay always jogs a mile seems a short distance to him'*, readers slow down at the word 'seems', as they must reinterpret the sentence (Frazier & Rayner, 1982). 'Jogs' occurs as a transitive verb more frequently than as an intransitive verb, 'a mile' is initially taken to be the object of the sentence. When reading the word 'seems', a transitive verb that requires a noun to be the subject, 'jogs' must be reanalyzed as a transitive sentence.

Such sentences highlight a central issue in sentence processing: is parsing serial or parallel (Gibson & Pearlmutter, 2000; Meng & Bader, 2000)? As this question is not central to the experiments in the dissertation, we will not got into detail on what evidence exists for each viewpoint. Instead, we wish to highlight the fact that such sentences show that syntactic structure for upcoming words is expected, as seen through increased processing difficulties when the expected structure is violated.

Grammatical context. Predictability effects have also been seen in how grammaticality affects processing in monolinguals; in other words, words are read faster when their syntactic category is expected. For example, Wright and Garrett (1984) found faster response times in a lexical decision task depending on what preceded the character string. Response times to the word 'batteries' (a noun) are shorter when it is preceded by a sentence fragment ending in a preposition (e.g., '*For now, the happy family lives with'*, as compared to a sentence fragment ending in a modal verb (e.g., *'If your bicycle is stolen, you must'*). Adding a noun to the first sentence fragment creates a grammatical sentence, the opposite is found in regards to the second sentence fragment. The word 'batteries' is unexpected in both sentences; though the first sentence is grammatical, it has a low cloze probability. Grammatical context is the main difference between the two phrases. The result was later replicated, and found to also be true in a naming task as well (West & Stanovich, 1986).

Prediction effects are also seen in sentences containing the word 'either'; encountering the word 'either' the reader/listener can expect a structure of two constituents joined by the word 'or'. In a self-paced reading study, participants were found to read 'or his father' faster when the prior sentence fragment was 'Sam wants either his mother'/'Sam either wants his mother' than 'Sam wants his mother'. The presence of 'either' was enough to facilitate reading a second constituent preceded by the word 'or' (Frazier & Clifton, 2001). A further study found that the facilitation effect due to the presence of the word ‘either’ carried into a spillover region after the second constituent of the word ‘or’ (e.g., In the sentence, ‘The team took either the train or the subway to get to the game’; the spillover region is ‘to get to the game’). Participants showed longer reading times and more eye movement regressions when the word ‘either’ was not present (Staub & Clifton, 2006).

Faster processing of words in grammatical contexts has been shown to be true even in minimal phrasal contexts, in that grammatical bigrams (sequences of two words) are read faster than ungrammatical bigrams (Vigliocco et al., 2008). This suggests that syntactic representations are at play immediately upon word recognition. For example, when reading ‘the’, a reader is most likely to expect a noun based on grammatical knowledge of English, so a noun like ‘book’ is more easily processed than a verb like ‘jumped’. This research demonstrates that syntactic predictability effects need not be embedded in sentences, but can occur in two word phrases.

The previous research was conducted in monolingual populations. In this dissertation we extend the concept of studies grammaticality effects into bilinguals.

Current Experiments

We are interested in examining word-by-word processing in bilinguals and language learners. Specifically, we wished to look at predictability effects caused by grammaticality across languages. As discussed previously, grammaticality effects have been found with language even in minimal grammatical contexts in monolinguals (Vigliocco et al., 2008); we were interested to see if the same holds true in two-word pairs when one word is in one language and the other in another.

Methodology

Lexical decision task. The experiments were carried out using a two-string lexical decision task, in which participants had to indicate whether or not both strings were words. A string was considered a word if it existed as such in either Spanish or English; participants were told that they should answer ‘yes’ if both were words, regardless of language, and ‘no’ if one or both strings were not words. Lexical decision tasks have long been used to investigate language processing, with faster reaction times associated with facilitated processing (Meyer & Schaneveldt, 1971; Ratcliff et al., 2004).

In our study, we were primarily interested in how grammaticality (whether or not the two words formed a grammatical constituent) affected processing. Instead of overall reaction times we looked at reading time of the second word, a more fine-grained measure; an assumption was made that grammaticality would not affect the reading time of the first word. First pass reading time was used as our dependent measure, as it has been associated with early processing and can be used to measure words of varying lengths (Clifton et al., 2007). Participants were required to focus on a point past the second word to prevent participants from leaving their eyes resting past processing while indicating the response. Additionally, participants were incentivized with a reward to prompt fast, accurate responses.

Procedure. Participants were shown a series of experimental and filler trials. All critical trials were 'yes' trials, and consisted of 'grammatical' trials and 'ungrammatical' trials. Grammatical trials consisted of a determiner-noun pair, and ungrammatical trials consisted of adverb-noun pairs. Reading time of the noun was always the dependent measure, which allowed us to eliminate variability due to syntactic category (Folk & Morris, 2003). To see if a grammaticality effect is present, we compare the reading time of a noun after a determiner (grammatical) as compared to after an adverb (ungrammatical), in word pairs with the same language combination. Faster reading times for the noun following a determiner as compared to the reading time of a noun following an adverb indicates a predictability effect due to grammaticality effect had been found.

Grammaticality was kept constant in both English and Spanish, so a word pair that forms a grammatical constituent in English does so as well in Spanish. This allowed us to determine the grammaticality of code-switching phrases, where a word pair that is grammatical in Spanish and English is also assumed to be grammatical in a combination of English and Spanish words.

Overview of Experiments

Experiments 1 and 3 were run on Spanish-English bilinguals recruited from the Ann Arbor area. Participants were tested on different stimuli, and proficiency measures were also different across experiments. In contrast, Experiments 1 and 2 shared the same stimuli and many of the same proficiency measures, but differed on participant type. Rather than looking at bilinguals, native speakers of English studying Spanish as an L2 at an advanced level were tested in Experiment 2.

Experiment 1: Syntactic category predictability in word pairs in Spanish-English bilinguals. In this study, we were primarily interested in whether or not Spanish-English bilinguals would show syntactic category predictability effects when grammaticality was kept constant. Stimuli consisted of words that learners often learn early when learning a language. Bilinguals were given a battery of proficiency tests for both English and Spanish, and language history information as well as attitudes towards code-switching information were collected after participants finished participating in the lexical decision task.

We expected to find grammatical predictability effects, as they have been found within language in studies on monolinguals. As we were looking at bilinguals, we did not expect to find an interaction between the language of the critical word and its grammaticality, because grammatical predictability effects should be observed in both English and Spanish. We were most interested in seeing whether or not grammatical predictability effects would be found in mixed language pairs, as this would indicate whether or not grammatical predictability was language-specific or language-independent. That is, the predicted grammaticality effect should not interact with 'congruency,' which is the factor distinguishing same language pairs from mixed language pairs. Finding an interaction with language congruency (suggesting that grammaticality effects are not found in mixed language pairs) would indicate that grammatical predictability is language-specific; not finding an interaction would indicate that grammatical predictability is language-independent.

Experiment 2: Syntactic category predictability in word pairs in Spanish L2 learners. In this experiment, we were interested as to whether L2 learners of Spanish would show the same predictability effects as bilinguals, or if grammaticality effects are linked to language 'native-ness' or moderated by proficiency. Advanced L2 learners of Spanish (all Spanish majors and minors studying Spanish in Spanish-only classrooms) were recruited at the University of Michigan. Stimuli in the lexical decision task was the same as Experiment 1, as were most of the proficiency tests/questionnaires (L2 learners were given additional proficiency measures for Spanish, and less for English).

As in Experiment 1, we expected to find a grammatical predictability effect, replicating Experiment 1. We were interested in seeing the effect of language native-ness on grammatical predictability; specifically, an interaction between the language of the critical word and its grammaticality would indicate that native-ness moderated the strength of the predictability effects, as it would indicate either a lack of or weaker predictability effects when the second word was in Spanish. As with bilinguals, we were also interested in seeing whether or not there would be an interaction between grammaticality and language congruency, indicating whether or not predictions were language-independent (no interaction) or language-specific (interaction). Furthermore, we were interested in seeing whether or not there would be an interaction between grammaticality, language, and language congruency, as it might indicate that grammatical predictability effects were only found in English-English word pairs.

Experiment 3: Conditional probabilities and predictability in Spanish-English bilinguals. In this experiment, we looked at bilinguals recruited from the Ann Arbor area (none of the same bilinguals who took part in Experiment 1 took part in Experiment 3). As in Experiments 1 and 2, participants completed a two-string lexical decision task with word pairs. No language proficiency measures were taken, nor language history/code-switching attitudes; only self-rated ability in Spanish and English. The stimuli used were different than in Experiments 1 and 2, having more variety in syntactic category and frequency (not necessarily covered the first semester of Spanish courses).

In this experiment, we wished to take into account that syntactic category can be ambiguous, as the same word can have more than one syntactic category (e.g., 'jump' can be a noun or a verb). This can lead to difficulties in labeling a word pair as 'grammatical' or 'ungrammatical'; due to the fact that det-V is ungrammatical and adv-V is grammatical. This ambiguity means that the many of the word pairs can have both a grammatical and an ungrammatical reading.

To account for category ambiguity, we did not use grammaticality as a categorical variable. Instead, we calculated two conditional probabilities, each indicating the grammatical category of one of the words of the pair and the exact letter string for the other word in the pair. These conditional probabilities gave a more granular view of grammaticality based on word probabilities. We then interpreted the data in a similar manner as we had in Experiments 1 and 2, substituting the conditional probabilities for grammaticality, and seeing how syntactic category expectation is affected by the conditional probabilities.

Overview of Dissertation

The following dissertation contains the experiments described above, and the analyses performed on the data collected from the experiments. It is organized in six chapters.

- Chapter 1: Introduction chapter which provides a wide overview of the topics which will be discussed as well as an overview of the experiments performed.
- Chapter 2: Experiment 1, *Syntactic category predictability in word pairs in Spanish-English bilinguals.* This chapter looks at the study conducted on Spanish-English bilinguals, and will show that evidence for language-independent syntactic category expectation was found.
- Chapter 3: Experiment 2, *Syntactic category predictability in word pairs in Spanish L2 learners.* This chapter looks at the study conducted on advanced Spanish L2 learners, and gives evidence

for the two major predictions, which are that language-independent syntactic category expectation is found and that it does not interact with Spanish proficiency.

- Chapter 4: This chapter compares the results of Experiments 1 and 2, examining the role of 'native-ness' in syntactic category expectation, and allowing us to look closer at whether or not bilinguals and language learners showed the same effects. Additionally, a closer look is taken at the proficiency measures, and how the different measures interact with grammatical predictability.
- Chapter 5: Experiment 3, *Conditional probabilities and predictability in Spanish-English bilinguals.* This chapter focuses on whether conditional probabilities are related to syntactic category expectation; specifically, greater expectation was found as one of the conditional probability measures (w2ccProb) increased.
- Chapter 6: General discussion, bringing together the results of all three studies as well as overall conclusions of the dissertation.

Chapter 2

Experiment 1: Spanish-English Bilinguals

Introduction

When processing speech, anticipating the next word leads to faster processing. Some words may be more predictable than others, for various reasons. Some possibilities are they may be linked semantically (Meyer & Schaneveldt, 1971), the words may frequently occur together (Monsell, 1991), there may be a common context that words may often share (McDonald & Shillcock, 2001), or the word may fit into the syntactic structure of the sentence being processed (Wright & Garrett, 1984; West & Stanovich, 1986). In regards to our study, we are most interested in seeing whether grammar-driven predictability effects are seen across languages.

In this introduction we will cover further information on code-switching, before explaining Experiment 1 and talking about how it relates to models of bilingual language representation.

Code-switching in Bilinguals

As we mentioned previously, bilinguals often code-switch, changing languages mid-sentence. What exactly constitutes a code-switch has been debated; does a sentence in one language except for one word constitute a code-switch or a borrowing? Is it a borrowing only if the phonetics are made to fit the main sentence? If a bilingual switches languages between sentences, is that a code-switch? Or is a change in languages only considered a code-switch when it occurs mid-sentence?

Part of the answer to these questions lies in what question one wants to answer regarding code-switching. For example, one common question investigated is why do bilinguals code-switch? In one study looking at Luyia speakers in Kenya, the most prominent reason was to change the social arena a conversation was taking place. A speaker would change from Luyia (the familiar, home language) to either English or Swahili to emphasize business, anger, or to signal a topic change. Most code-switches in this study were seen between sentences (Myers Scotton & Ury, 1977). In contrast, a study which focused on one-word switches found that code-switches most often took place when the speaker either could not recall the word in the language they were primarily using, or if a word encompassing a specific concept did not have an exact translational equivalent (Heredia & Altarriba, 2001).

In this dissertation, we focus on how languages interact during word by word processing. We will use the term code-switch to refer to language changes that occur mid-sentence, i.e., intrasentential code-switches. Given that we are not focusing on the social functions of code-switching, we will look at models of code-switching focusing on the mechanics of how it takes place.

Poplack's constraint-based model of code-switching. After observing a community of Spanish-English bilinguals (characterized by Puerto Rican immigrants and those descended from immigrants living in New York City), Shana Poplack observed several patterns among the mixed language utterances (1980). The patterns seemed to be focused on two simple constraints: that of equivalence, and that of free morpheme (Sankoff & Poplack, 1981).

Equivalence constraint. The equivalence constraint states that the surface structures must be the same for code-switching to occur. In other words, if the grammatical structure is not shared across languages, the switch will not occur.

Free morpheme constraint. The free morpheme constraint states that a switch may not occur with a bound morpheme and a lexical form unless the lexical form has been integrated phonetically into the language of the bound morpheme. For example, one would not see the English word 'speak' being code-switched with the bound verb root 'habl' and attached to the morpeme –ando (gerund morpheme in Spanish), forming something like 'speakando'. This is because 'speak' has not been phonetically integrated into Spanish, easily noticed because it violates Spanish phonetics by starting with a constant cluster beginning with 's'. If 'speak' had been phonetically integrated, it could be bound to '-ando', however, it could not be considered a borrowing rather than a code-switch.

Problems with model. The Poplack model cannot explain why certain code-switches which do not violate these two constraints are found to be unacceptable. For example, code switches that occur mid-periphrastic construction, (e.g. 'they had ido al cine' [they had gone to the movies]) are not accepted by Spanish-English bilinguals, even though the word order is preserved and the code-switched segment was not a bound morpheme (Toribio, 2001a).

Additionally, researchers looking at more dissimilar languages than English and Spanish have found code-switches that violate the equivalence constraint, with code-switches occurring in cases when the surface structure is different across languages. This has been seen in diverse language pairings, such as Welsh and English (Deuchar, 2005), Spanish and Hebrew (Berk-Seligson, 1986), and French and Moroccan Arabic (Bentahila & Davies, 1983). In some language pairs, bilinguals accept code-switches which do not follow the surface order of either language. French-Moroccan Arabic speakers will often switch determiner-noun pairs instead of simply the noun following some Arabic determiners, forming an order of det-det-N, ungrammatical in both languages (Bentahila & Davies, 1992). The free morpheme constraint has also been found to be violated in some language pairs; for example, Spanish-Nahuatl bilinguals produce words consisting of Spanish verbs attached to Nahuatl affixes (MacSwan, 2000).

Other constraint-based models. Poplack's model was the first in a class of models that looked for grammatical constraints governing when it was possible to make a code-switch. Other researchers proposed different constraints, a few of which are described below.

Closed-class constraint. One such model proposed after observing Marathi-English bilinguals theorized a closed-class constraint; closed-class items such as determiners, prepositions, etc. can not be switched (Joshi, 1985). However, looking a different language pairs found violations of that constraint, such as Farsi-English (Mahootian & Santorini, 1996) and Italian-French (Di Sciullo et al., 1986).

Functional head constraint. Another model instead states that the language of the complement *f*-selected by a functional head and that of the functional head must match, like any other linguistic feature (Belazi et al., 1994). Such a constraint would rule out switches between determiners and NPs, which have been found to be the most common type of code-switch in English-Spanish (Poplack, 1980).

Constraint-free model. Constraint-based models of code-switching face the problem that they are based on certain language pairs (whether they are one or several); as soon as a new constraint is proposed, other researchers test the constraint on other data from other languages and find cases where the constraints are violated (Gardner-Chloros & Edwards, 2004). As an alternative, a constraint-free model of code-switching has been proposed, which states that there are no universal constraints on code-switching (MacSwan, 1999; MacSwan, 2000). Instead, any constraints come from the mixed grammar formed by the particular language pair being looked at, driven by the properties of each of the languages.

Code-switching in the experiment. In the experiments covered in this dissertation, we looked at Spanish-English bilinguals (and language learners), and so materials were designed primarily as following the equivalence constraint, as that was most relevant for our purposes. (No words that could be part of one language or the other were used, nor any strings violating the free-morpheme constraint.) Grammaticality was kept constant in all two word phrases, complying with the constraint. We considered det-N mixed word pairs to be grammatical as they are the most common type of code-switch in Spanish-English code-switching; as such, we are ignoring the functional head constraint. As we looked only at two-word pairs, not enough information is present to determine or not the closed-class constraint is violated.

We were not looking at social influences in code-switching; however, as a control measure, we collected data on language use and code-switching so we could look at whether code-switching use affects processing of mixed language word pairs. Furthermore, we collected data on code-switching phrases to see whether or not awareness of specific Spanish-English code-switching constraints affected processing.

Experiment

We were interested in looking at grammatical predictability across languages. The question of predictability becomes more complex when we look at bilinguals, since the nature of what is shared between languages is still under debate. Our main question of interest is whether shared grammatical representations can lead to predictability effects across languages during online processing.

We looked at grammatical predictability using a two string lexical decision task using Spanish-English bilingual adult participants, during which we looked at how reading was affected by different types of word pairs. Word pairs were used to control grammaticality more finely; specifically, to reduce potential noise from both incidental grammatical combinations and either list reading effects when words were not in a grammatical phrase.

Participants were shown two letter strings and asked to judge whether or not the two strings formed words, in either Spanish or English. We looked at word pairs which formed a grammatical phrase in both English and Spanish, determiner-noun, e.g., ‘any car’, and word pairs which did not form a grammatical phrase in either language, adverb-noun, e.g., ‘slowly cat’. Word pairs occurred in all language combinations, leading to same language pairs, English-English and Spanish-Spanish, as well as mixed language word pairs, Spanish-English and English-Spanish. Participants were eye-tracked while reading these bigrams to allow us to assess how quickly the noun was read.

Hypotheses. Participants were expected to show predictability effects of grammaticality in same language word pairs, due to the fact that earlier research has shown that words are read faster in a grammatical constituent as compared to otherwise, regardless of the frequency of the two words occurring together (Wright & Garrett, 1984; Clifton et al., 2007). Predictability effects would be observed as participants being quicker to read the noun when it was following a determiner as compared to it following an adverb. We predicted to see this in both English-English and Spanish-Spanish word pairs, as our participants are fluent in both Spanish and English. Our main question of interest was whether the same pattern would hold through in mixed language bigrams.

Grammatical predictability effects not found across languages. If no grammaticality effects were found in mixed language pairs, it would suggest that languages have separate grammars, supporting models of bilingualism such as de Bot's bilingual speaking model (1992). While primarily a production model, it holds that bilinguals have different formulators for separate languages they also have separate comprehenders for each language. In this case, predictability effects should not be seen across languages if they are not connected. Though the de Bot states that the model does not adequately explain code-switches, he allowed that the different language formulators might be connected in some way to describe code-switches. However, the connections across languages would be weaker, as can be seen by language switch costs and the fact that code-switches are uncommon compared to total language production. So too might the comprehenders be; however, in this case, we would expect to see less grammatical predictability effects in mixed language pairs, since the connections would be weaker as compared to same language pairs.

Grammatical predictability effects found across languages. If instead grammaticality effects were found in the mixed language pairs, it would instead support a shared syntax model, with grammar being shared across languages. This is due to the fact that the first word would predict words that would fit possible grammatical structures, with activation spreading from the combinatorial nodes to the lemma modes (Figure 2). Since language nodes would be activated after the lemma nodes, activation should be equal regardless of the language of the lemma. By this interpretation, we would expect that grammatical predictability effects found in mixed language pairs should be of equal strength as to those found in same language pairs.

Other expected findings. Though our main question is regarding grammatical predictability, we also expect to find several other patterns in our data. In accordance to known effects found in reading, we expect to see more frequent words read faster than less frequent words, as well as shorter words read faster than longer words. Spillover effects are also likely, as both the frequency and length of the previous word have been found to affect the reading time of the current word.

We expected that participants who are more fluent in Spanish will read Spanish faster than participants less fluent in Spanish. We also expected a language switch cost, with participants showing longer reading times for words in a mixed language pair as compared to words not in a same language pair. A practice effect was considered likely, as participants often show practice effects in cognitive tasks, getting better at the study as they go on—reading times were expected to decrease slightly as participants grew more used to the task.

Method

Participants

Fifty fluent bilingual Spanish-English speakers from the University of Michigan participated in the experiment. Participants received $25 base pay as well as bonuses based on speed and accuracy, for a total average payment of $32.62 (range: $29.14 : $35.30, sd = $1.30).

The mean age of the participants was 23.53 years (range: 18 : 63). Most participants (n = 30) were children of Spanish-speaking immigrants, and said they had learned both languages simultaneously as infants (n = 24) or had learned Spanish in infancy and English once they started daycare/pre-school (n = 9). Other participants (n = 8) reported having learned English in early childhood (between the ages of 5-12), and some (n = 9) learned English as a second language in adulthood, though all participants lived in the United States at least one year and reported speaking English for at least 5 years.

Materials

Word stimuli. Participants saw English and Spanish words and nonwords during the study. All critical trials consisted of two words, so we will describe the word stimuli before turning to the nonword stimuli. Due to the fact that this study was designed to also be used with Spanish L2 learners, Spanish words used in the study were drawn from introductory Spanish textbooks [Aventuras, 3th edition (Donley, 2010); Dos Mundos, 6th edition (Terrell et al., 2005)] used in the first year of teaching as much as possible. This includes all words except for 61 adverbs (out of 172 total adverbs used in experiment), because not enough adverbs were present in the books. The additional 61 adverbs were the most frequent adverbs as listed in the Corpus del Español (Davies, 2002) that were not in the books.

English words were translations of Spanish words whenever possible. Different participants saw all language permutations of a word pair; for example, if one participant saw the English-English adverb-noun word pair ‘mutually elevators’, another participant would see the English-Spanish version ‘mutually ascensores’, another the Spanish-English version ‘mutuamente elevators’, and another the Spanish-Spanish version ‘mutuamente ascensores’. Due to the fact that there is not one-to-one correspondence for all words, especially function words, in some cases words were matched by syntactic category and grammatical number instead of being translations. Participants saw most words once in the study (in either English or Spanish), the exception being determiners, all of which appeared 1-2 times in each language.

English word frequencies (per million) were drawn from the Corpus of Contemporary American English (Davies, 2008), and Spanish word frequencies (per million) from the Corpus del Español. Overall, because Spanish words have gender, English words tended to be more frequent (average log frequency = 3.9) than Spanish words (average log frequency = 3.0). This difference in frequency was seen in all syntactic categories where words have gender in Spanish: determiners, adjectives, nouns, and pronouns. (Average log frequency for Spanish word with gender = 3.1; average log frequency for English words whose Spanish equivalent has gender = 3.7) This frequency difference was not seen in syntactic categories where words do not have gender in Spanish: adverbs, prepositions, conjunctions. (Average log frequency for Spanish word without gender = 5.6; average log frequency for English words whose Spanish equivalent does not have gender = 5.5) Verbs also showed a difference in frequencies across languages; Spanish verbs have more conjugated forms than English verbs, leading to English verbs being more frequent. (Average log frequency of Spanish verbs: 0.9; average log frequency of English verbs: 3.5)

Nonwords. Nonwords were created to follow the phonotactic rules of the respective languages, by random selection of words that were not used in the study and then deleting, adding, or changing one or two characters, avoiding homophony with real words. All words were checked in the online Merriam-Webster English dictionary (http://www.merriam-webster.com/) and the online dictionary produced by the Real Academia Española (http://dle.rae.es/?w=diccionario) to make sure they did not exist in either

English or Spanish. In total, there were 100 English nonwords and 100 Spanish nonwords; each participant saw each nonword once during the experiment.

Word pairs. Both critical and filler word pairs were classified as either grammatical or ungrammatical. Grammatical word pairs fell into one of the following categories: determiner-noun, adverb-verb, adjective-adjective, noun-verb, and verb-noun. Ungrammatical word pairs were of the following categories: adverb-noun, determiner-determiner, determiner-preposition, determiner-conjunction, determiner-pronoun, determiner-interjection, noun-determiner, noun-adverb, or noun-noun. In total, in the 'yes' trials, each participant saw 152 grammatical word pairs and 148 ungrammatical word pairs. Critical word pairs were equally divided across language combinations, so each participant saw 18 grammatical word trials (determiner-noun) as English-English, 18 as English-Spanish, 18 as Spanish-English, and 18 as Spanish-Spanish. Equal numbers of ungrammatical word trials (adverb-noun) were shown for each language combination. Each participant also saw twenty grammatical filler trials and nineteen ungrammatical filler trials per language combination. In addition, each participant saw 200 'no' trials, each containing one nonword. (See Table 1 for a breakdown of the types of trials in the experiment.)

Table 1: Types of trials in Experiment 1.

Grammatical Critical Trials	72 determiner-noun
Ungrammatical Critical Trials	72 adverb-noun
Grammatical Filler Trials	36 adverb-verb 12 adjective-adjective 16 noun-verb 16 verb-noun
Ungrammatical Filler Trials	36 determiner-*function* word determiner-determiner/determiner-preposition/ determiner-conjunction/determiner-pronoun 16 noun-determiner 16 noun-adverb

	8 noun-noun
'No' Filler trials	36 determiner-nonword 36 adverb-nonword 12 adjective-nonword 12 verb-nonword 8 noun-nonword 4 *function* word-nonword preposition-nonword/conjunction-nonword/pronoun-nonword 16 nonword-adverb 16 nonword-determiner 20 nonword-adjective 20 nonword-noun 20 nonword-verb 8 nonword-*function word* nonword-preposition/nonword-conjunction/nonword-pronoun

Critical trials were chosen to be determiner-noun because it is the most common syntactic bigram in both English and Spanish, as well as the most common point for a code-switch in English-Spanish code-switching sentences (Liceras et al., 2008). Ungrammatical critical word pairs were adverb-N, which does not occur in grammatical constituents of either language. (It can occur across constituents, e.g. *After diving deeply, whales eat well.*)

Word pairs were created by first separating the word stimuli by syntactic category into separate lists, and then randomly ordering each list (e.g., nouns, verbs, 'function words', etc.). Then, for each type of word pair, the necessary number of words were taken from appropriate syntactic list, paired together, and then controlled to match for gender and number (if there was a mismatch, one of the words would be switched with another pair to prevent it). Word lists were divided so as to have equal numbers of language-language combinations, equally spread through each type of word pair. For simplicity, we will refer to total numbers of word pair types shown rather than specifying the number per language-combination. (Table 2 shows critical word pairs.) There were a total of 500 trials, 300 yes trials and 200 no (nonword) trials. Among the yes trials, there were 152 grammatical trials and 148 ungrammatical trials.

Eight different testing lists were formed, so as to prevent specific order or word pair effects. Lists 1-4 vary only by language; word pairs are otherwise in the same order. Lists 5-8 are composed of the same words as Lists 1-4, but with words in different word pairs, and the order of the type of word pairs was different. Once all word pairs were made, the list was randomly ordered. To clarify, if a word pair was shown with Word 1 in English and Word 2 in English in List 1, then that same word pair was shown with Word 1 in English and Word 2 in Spanish in List 2, with Word 1 in Spanish and Word 2 in English in List 3, and with Word 1 in Spanish and Word 2 in Spanish in List 4. Lists 5 through 8 were formed independently to Lists 1-4; they varied by language the same way, but were formed of different word pairs made of the same words used to make up Lists 1-4.

Table 2: Examples of word pairs in Experiment 1.

	Grammatical	Ungrammatical
Same language	my dogs la comida	strongly amusement prácticamente ballenas
Mixed language	some conciertos dieciocho feathers	propiamente nationality after maestra

Language assessments and questionnaires. To assess Spanish proficiency, participants were given the lexTALE-ESP. The lexTALE-ESP is a Spanish lexical decision task with 90 trials (Izura et al., 2004). During the task, participants must decide whether the word presented is a Spanish word or not; words of different frequencies are tested so as to avoid floor (all wrong) and ceiling (all correct) effects. Ultimately, it is a vocabulary-based proficiency measure; in previous work, it has shown different scores for high proficiency and low proficiency language learners, as well as language learners and native speakers. In our study, the test was changed into a Qualtrics format so as to be easier for participants to complete. Words that were tested in the lexTALE-ESP were not used in the eye-tracking task.

Equivalently, as a control of English proficiency, participants were given the lexTALE (Lemhöfer & Broersma, 2012). The lexTALE is an English lexical decision task with 60 trials, looking at words of different frequencies (the lexTALE-ESP is a Spanish version based on it, but uses different words and has more trials). Similar to the lexTALE-ESP, it is a vocabulary-based measure of proficiency. As with the lexTALE-ESP, words that appeared in the lexTALE were not used in the eye-tracking task. The lexTALE was also placed onto Qualtrics to be easier for participants to complete.

To gather information on participant's language background and code-switching attitudes, participants were given two questionnaires, both presented on the computer: the ACSES, designed to look at code-switching behavior and attitudes in Spanish-English (Escobales, 2014) and the LHQ 2.0 (language history questionnaire), that asks questions solely about language history and how and where each language was acquired, as well as how each language tends to be used (Li et al., 2014). The main questions had to do with the amount of time the participant spent code-switching, as well as what attitudes the participant had regarding code-switching in general.

Participants were also given a sentence assessment task where participants were asked to rate 10 sentences containing an intrasentential code-switch (5 from English to Spanish, 5 from Spanish to English). Sentences were broken down into 5 acceptable code-switches, and 5 unacceptable code-switches. Although this test was developed by the author, 'acceptable' and 'unacceptable' code-switches were taken from studies that looked at corpora of code-switching speech and studies that tested both bilinguals and language learners on sentences that contained a code-switch that either occurred naturally or not (Belazi et al., 1994; Toribio, 2001a; Toribio, 2001b). Sentences were tested on a 7-pt Likert scale, with participants asked to judge how acceptable the sentence was to them.

As a final measure of proficiency, participants were also tested on an elicited imitation task (EIT) in both English and Spanish. During the EIT, participants are asked to repeat sentences that they hear.

Unlike with the lexTALE and lexTALE-ESP, which are explicit tasks—participants tested directly on linguistic knowledge; the EIT is an implicit task, based on a participant's ability to comprehend, remember, and reproduce a sentence in the language tested. While under debate if it is the best measure of global proficiency (Sarandi, 2015), it is likely testing different aspects of linguistic knowledge than vocabulary-based measures, as that is primarily a recognition task, with neither a grammatical component nor a speaking fluidity assessment. In this study, we primarily used the lexTALE and lexTALE-ESP as our measures of proficiency, however, EIT data was collected as a secondary measure of proficiency.

The Spanish EIT used was developed by Harriet Bowden (2016), based on principles developed by the originators of the task (Ortega et al., 1999). During the task, participants were recorded as they repeated 30 Spanish sentences that they heard through headphones. Sentences ranged from 7 to 17 syllables and had a variety of grammatical structures. Prior to the Spanish sentences, participants repeat 6 English sentences ranging from 7 to 17 syllables. Experimental recordings heard by participants were made by a female native Spanish speaker from Ecuador in the Bowden lab. The Spanish EIT was administrated using Praat.

We developed the English EIT to mirror the structure of the Spanish EIT. Participants were recorded as they repeated 30 English sentences through headphones, following a practice where they repeated 6 Spanish sentences ranging from 7 to 17 syllables. Sentences ranged from 8 to 19 syllables, and had a variety of grammatical structures. Experimental recordings were made by a female native English speaker from South-East Michigan in the Boland lab. As with the Spanish EIT, the English EIT was administrated using Praat.

Procedure

Participants took part in a two hour-long session, composed of one hour of a bigram lexical decision task followed by one hour of language assessments and questionnaires. Participants completed a two-string lexical decision task (Meyer & Schaneveldt, 1971) while being eye-tracked using an Eyelink 1000 eye-tracker set 2 feet from the screen as measured from the chinrest. Before being eye-tracked, participants were calibrated using the Eyelink calibration software.

Each trial started with a single point drift correction in the center of the screen. Participants then focused on a fixation cross on the left side of the screen which automatically triggered the appearance of a two-word list followed by another fixation cross, all in 16-Pt Courier font (Figure 3). Participants were instructed to view both words, fixate on the rightmost cross, and indicate their response using the keyboard. Participants pressed a button indicating 'yes' if *both words* existed in either English or Spanish, and a different button otherwise. Instructions were presented in English, and the experimenter answered clarification questions after presenting the instructions. Responses were only accepted while fixated on the rightmost cross. The crosses were included to promote natural left to right reading patterns and to make it more likely that gaze duration on each word was due to reading, not making the lexical decision.

Figure 3: Example of a mixed language determiner-noun trial.

Participants received points for correct answers and lost points for incorrect answers, following the formula given in Figure 4, prompting participants to follow a consistent speed-accuracy tradeoff. Feedback was given after each trial, stating correctness, time taken to answer, and points gained or lost. Every twenty-five trials participants were shown how many total points they had accumulated during the course of the experiment. Points were converted to dollars by dividing by 1000; so if a participant finished with a total of 6000 points, they would receive a bonus of $6.00.

$$Bonus\ for\ correct\ answers = \frac{4500 - reaction\ time\ (ms)}{150}$$

$$Penalty\ for\ incorrect\ answers = 50 - \frac{4500 - reaction\ time\ (ms)}{150}$$

Figure 4: Formula for calculating reward/penalty in Experiment 1. If a participant takes 1050ms and answers correctly, the bonus would be 23pts. If the participant answers wrong at 1050ms, the penalty would be -27pt. If reaction time is greater than 4500ms, the bonus is 0pt, and penalty is 50pts.

Participants completed a total of 500 trials, preceded by 8 practice trials. Each participant was randomly assigned one of eight stimulus lists for a total of six or seven participants per list. After completing the bigram lexical decision task, participants completed the language assessments and surveys in the following order: lexTALE-ESP, lexTALE, code-switching sentence judgment task, ACSES, LHQ 2.0, Spanish EIT, English EIT. Unlike in the bigram lexical decision task, no feedback was given following the assessments.

Scoring

Language proficiency measures.

lexTALE and lexTALE-ESP. Both the lexTALE and lexTALE-ESP were rated as described by the original authors, with a total percentage correct calculated (participants had to correctly identify words and reject nonwords). In the lexTALE, participants had to correctly identify 40 words and 20 nonwords; in the lexTALE-ESP, participants had to correctly identify 60 words and 30 nonwords.

Code-switching sentence ratings. Participants were each given two scores for the task: 'grammatical' sentence score, and an 'ungrammatical' sentence score. The two scores were calculated by taking the mean of the rating given by the participant on the 5 grammatical and 5 ungrammatical sentences, respectively. Ratings were on a 1-7 Likert scale, with 1 as unacceptable, and 7 being completely acceptable.

ACSES and LHQ 2.0. Both measures numerous sources of data; currently, we have 6 values calculated from the information gathered. From the ACSES, we have the mean self-rating for Spanish and English (calculated from speaking, comprehension, reading, and writing values—question 5), mean self-rating of percentage of times spent code-switching (question 7), mean time spent code-switching with the six people you talk to the most (question 10); and mean positivity rating (question 24). From the LHQ 2.0, we have data about which language a participant speaks natively.

Spanish EIT and English EIT. The data from 2 participants' Spanish EIT and 1 participant's English EIT are missing due to recording issues, leading those 3 participants to be discarded from all analyses for the EIT. The other 47 participants had data for both Spanish and English EIT, and their data was used for the analyses.

Each sentence was scored on a value of 0 (no information repeated) to 4 (sentence perfectly formed); mid-values show degrees of success in repeating the sentence. All subjects were rated by the same rater, and a different rater checked 20% of the participants. Ratings between the raters were shown to be correlated to a high degree $r(13) = 0.95$, $p < .001$, showing that rating was consistent.

Lexical decision task. For our analyses, we looked only at critical grammatical and ungrammatical trials. Filler trials and 'No' trials were not analyzed. Due to the fact that mean reading times varied by participant (range: 235ms : 604ms), outliers were calculated by participant. Words with reading times less than 50 ms or more than three standard deviations from that participant's mean reading time were dropped from analysis (0.94% of trials), leaving 7132 total trials used in the analyses.

We performed several sets of analyses. We first conducted an accuracy analysis. For the accuracy analysis, we calculated each participant's average percentage correct (APC) by language x congruency x grammaticality, for a total of eight values per participant. Each percentage was calculated by dividing the number of correct trials by the total number of correct and incorrect trials in the particular language x congruency x grammaticality condition (trials that had been deemed outliers were not included in the calculation). In total, there were 400 values taken into the model, 8 per participant.

We also analyzed the first pass reading time of the second word as our primary measure. Only correct 'yes' grammatical and ungrammatical trials were analyzed, lowering the total number of trials analyzed to 6874. We also analyzed the probability of whether or not a participant made an eye movement regression from the second word back to the first word; as in the reading time analysis, only correct trials were scored. Trials that included an eye movement regression were scored as a 'hit', with a value of '1', and trials without an eye movement regression were scored as a 'miss', with a value of '0'. A final analysis looked at the first pass reading time of trials without regressions, lowering the total number of trials to 6223.

Results

Language Proficiency Measures

We looked at a total of 57 correlations and planned comparisons, so a Bonferroni correction was applied. Probability values had to be less than $8.77*10^{-4}$ to be considered significant. Table 3 shows the average score for each measure.

Table 3: Language proficiency and use means in Experiment 1.

lexTALE score	**90.25%**
English self-rating [1:7]	**6.36**
English EIT [1:4]	**3.71**
lexTALE-ESP score	**79.92%**
Spanish self-rating [1:7]	**6.20**
Spanish EIT [1:4]	**3.78**
Estimated Percent Time Code-switching	**23.68%**
Percent Time Code-switching with Top Conversational Partners	**35.73%**
Code-switching Positivity Rating [1:7]	**4.23**
Grammatical Code-switching Sentence Rating	**5.29**
Ungrammatical Code-switching Sentence Rating	**3.63**

ACSES and LHQ 2.0. Across all 50 participants, the mean self-rating for English was 6.36 (range: 3.5 : 7, sd = 0.78), and the mean self-rating for Spanish was 6.20 (range: 3.5 : 7, sd = 0.86). There was no significant difference in a participant's self-rating of English and Spanish.

The mean self-rating for the percentage of time spent code-switching was 23.68% (range: 0% : 90%, sd = 22.71%). The estimated percent mean time spent code-switching score with top conversational partners was 35.73% (range: 0% : 100%, sd = 24.76%). The two scores were positively correlated, $r(48)$ =

0.54, $p = 6.01*10^{-5}$, with a participant's self rating of time spent code-switching increasing with the percentage of time spent code-switching with their top conversational partners.

The two code-switching scores were however significantly different, $t(49) = -3.72$, $p = 5.19*10^{-4}$, with participants indicating that they spent more time code-switching with their top conversational partners than they rated themselves as having done so in the last week. This finding is not necessarily contradictory, as participants might have spoken to many people without code-switching, lowering the amount of time spent code-switching, but not lowering the mean amount of time spent code-switching with their top conversational partners.

Participants were generally neutral towards code-switching, giving a mean score of 4.23 (range: 1 : 7, sd = 1.52). There was a trend for code-switch positivity scores to be correlated positively with measures of time spent code-switching (with self-rating of time spent code-switching: $r(48) = 0.45$, $p = 1.14*10^{-4}$; with percentage of time spent code-switching with top conversational partners: $r(48) = 0.44$, $p = 1.08*10^{-4}$), indicating that participants who felt more positive towards code-switching spending more time code-switching in general.

Most participants spoke Spanish as their first language, with 35 indicating that they had not learned English until approximately age 4-5. Eight participants said they spoke English as their first language, and 7 participants indicated that they learned both languages concurrently since infancy. Though the sample sizes are too different to do any meaningful test, participants who spoke English first had the highest average positivity score, 5.15, participants who learned both languages concurrently having a mean score of 4.87, and participants who spoke only Spanish first having the lowest average positivity score, 3.89.

lexTALE and lexTALE-ESP. The lexTALE and lexTALE-ESP were used as our measures of basic proficiency during the task, as they are similar proficiency tests of Spanish and English which all participants completed. Across all 50 participants, the average score on the lexTALE was 90.25% (range: 61.25% : 100%, sd = 8.91%), and the average score on the lexTALE-ESP was 79.92% (range: 57.50% : 100%, sd = 10.98%). Participants scored significantly higher in the lexTALE (English) than the lexTALE-ESP, $t(49) = 4.61$, $p = 2.91*10^{-5}$, indicating that they were more proficient in English than Spanish. This is not surprising, as most participants were educated in primarily in English, though their first language might have been Spanish.

The lexTALE was positively correlated with participants' self-rating of English, $r(48) = 0.59$, $p = 7.89*10^{-6}$, indicating that participants were generally accurate in their self-rating. Similarly, lexTALE-ESP scores were positively correlated with participants' self-rating of Spanish, $r(48) = 0.66$, $p = 1.74*10^{-7}$. Neither lexTALE nor lexTALE-ESP scores were correlated with either code-switching use measure, nor with code-switching positivity scores.

Code-switching sentence ratings. Across all participants, the average grammatical sentence score was 5.29 (range: 1 : 7, sd = 1.83), and the average ungrammatical sentence score was 3.63 (range: 1 : 6.8, sd = 1.65). Ratings between grammatical and ungrammatical code-switching sentences were correlated, $r(48) = 0.67$, $p = 9.06*10^{-8}$, meaning that participants who rated grammatical sentences higher also tended to rate ungrammatical sentences higher. Participants rated grammatical code-switching sentences higher than ungrammatical ones, $t(49) = 8.64$, $p = 2.01*10^{-11}$, showing they accepted sentences conforming to known Spanish-English constraints as more grammatical than ones which violated those constraints. Sentence judgment scores were not correlated with either lexTALE or lexTALE-ESP scores, nor with self-reported measures of proficiency, code-switching percentage estimates, nor with code-switching positivity.

Spanish EIT and English EIT. The Spanish EIT and the English EIT were used as our secondary measures of proficiency. They were collected as a secondary measure to ensure that if the lexTALE and lexTALE-ESP were at ceiling level, a different measure of proficiency was available for comparison. For the 47 participants for which we had both scores, the average score (out of a maximum of 4) on the Spanish EIT was 3.78 (range: 2.90: 4.00, sd = 0.26), indicating high Spanish proficiency. The average score on the English EIT was 3.71 (range: 3.03 : 3.97, sd = 0.22), which is likewise high. Participants scored equally on both tests, $t(46) = 0.08$, $p = 0.94$.

The Spanish EIT score was correlated with the lexTALE-ESP score, $r(45) = 0.71$, $p = 2.14*10^{-8}$, and the English EIT score was correlated with lexTALE score, $r(45) = 0.55$, $p = 5.77*10^{-5}$. As the English EIT score and the Spanish EIT score were correlated with the lexTALE and lexTALE-ESP, respectively, we did not use the EIT scores in any further analyses. We report them to compare to other published studies and to show that oral fluency was also measured, and was correlated with vocabulary in each language. Spanish EIT scores were also correlated with self-reported measures of Spanish linguistic ability, $r(45) = 0.63$, $p = 2.49*10^{-6}$, but English EIT scores were not correlated with self-reported measures of English linguistic ability.

EIT scores were not correlated to code-switching sentence judgment scores, nor with any self-reported measure of code-switching or the code-switching positivity score.

Lexical Decision Task

Accuracy analysis. Average accuracy was 96.38% for critical trials, with highest accuracy for English-English grammatical trials (98.55%), and lowest for English-Spanish ungrammatical trials (94.36%). Due to very high rates of accuracy (Table 4), we were unable to fit a full parsimonious logistic regression model that accurately represented the data. We instead fit a linear mixed-effects model to the APC using the lme4 package in R (Bates et al., 2015; R Core Team, 2015). The model included fixed variables for grammaticality, language of the second word, whether or not the trial was a same or mixed language trial (congruency), all possible interactions between these three variables, and a random variable for subject. Variables were centered before being added to the model.

Proficiency as measured by the lexTALE and the lexTALE-ESP tests were included as fixed variables to take into account individual differences in English and Spanish proficiency, respectively. Proficiency was centered before being added to the model. Since we did not look at accuracy for each specific trial, but instead looked at each participant's average accuracy score across all trials as divided by grammaticality, language of the second word, and language congruency, we did not take into account any specific trial characteristics as factors, nor any factor to account for time during the experiment.

We followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. In other words, we first created a model that included all the predictors in the random effect structure, and then we looked at what proportion of the variance each variable accounted for, excluding the interactions. If it accounted for less than 1 percent of the variance, that predictor was removed from the next model. What proportion of the variance each predictor accounted for in that model was looked at, and again the same exclusionary criteria were used, until a final, convergent model was obtained with each variable in the random effect structure accounting for at least 1 percent of the variance. The final model included for subject, a random intercept, and slopes for the language of the second word and language congruency (Figure 5).

```
Model <- glmer(APC ~
               EnglishProficiency +
               SpanishProficiency +
               WordLanguage*LanguageCongruence*Grammaticality +
               (1 |subject) +
                 (0 + LanguageCongruence |subject) +
                 (0 + WordLanguage |subject) +
               data=bilingual, REML = F)
```

Figure 5: Linear mixed-effects model for accuracy analysis in Experiment 1.

No main effect of grammaticality was found, though a three-way interaction was found between language, congruency, and grammaticality), β = -0.048, $t(387) = -2.677$, $p = 0.008$. This three-way interaction is likely spurious, the result of grammatical trials showing a 3% higher accuracy than ungrammatical trials in English-Spanish word pairs (see Table 4). Accuracy is high for all conditions, with the accuracy values similar in all four conditions, and close to the mean accuracy overall (96.38%).

Same language trials were answered more accurately than mixed language trials, β = -0.018, $t(387) = -3.675$, $p < .001$, showing a language congruency effect. No main effect of language was seen, but there was an interaction between congruency and the language of the second word, β = 0.023, $t(387) = 2.562$, $p = 0.01$, driven by language switch cost when the second word was in English. This is likely due to the fact that participants were stronger in English than Spanish, and so English-English trials were easier than Spanish-English trials. A similar drop-off was not seen following a language switch when the second word was in Spanish for the same reason—because Spanish-Spanish trials are already challenging, the switch cost is partially mitigated by the fact that the first word is in English (Figure 6).

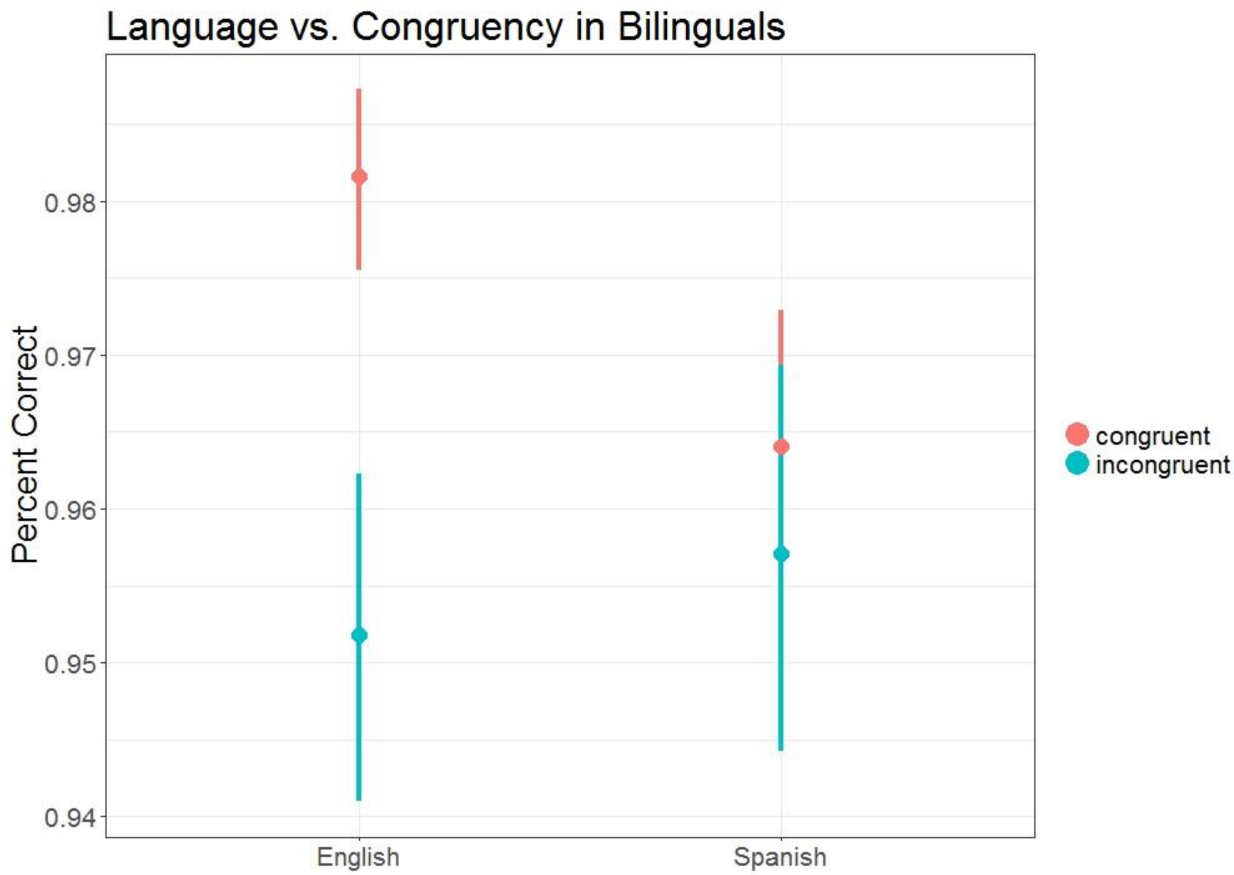


Figure 6: Congruency x language interaction for accuracy analysis in Experiment 1. Language refers to the language of the second word; error bars represent the standard error of the mean.

Of the language proficiency variables, Spanish proficiency was found to be significant, $\beta = 0.014$, $t(387) = 5.164$, $p < .001$, with participants being generally more accurate as their proficiency increased. There was a trend of English proficiency, $\beta = 0.005$, $t(387) = 1.791$, $p = 0.08$, with participants being generally more accurate as their proficiency increased, but it did not reach significance.

Table 4 shows accuracy rates by grammaticality, language condition, and congruency. For each row, the topmost number is the overall percentage correct per condition (not separated by grammaticality), followed by the percentage correct as divided by grammaticality (det-N is grammatical, adv-N is ungrammatical).

Table 4: Average accuracy rates in Experiment 1.

	Percent Correct	
English-English	98.16%	
	det-N	98.55%
	adv-N	97.77%
Spanish-English	95.18%	
	det-N	94.67%
	adv-N	95.70%
English-Spanish	95.77%	
	det-N	97.18%
	adv-N	94.36%
Spanish-Spanish	96.41%	
	det-N	96.33%
	adv-N	96.49%
Same Language	97.29%	
	det-N	97.44%
	adv-N	97.13%
Mixed Language	95.47%	
	det-N	95.91%
	adv-N	95.03%
Word 2 (English)	96.67%	
	det-N	96.60%
	adv-N	96.74%
Word 2 (Spanish)	96.09%	
	det-N	96.76%
	adv-N	95.42%
ALL	96.38%	
	det-N	96.68%
	adv-N	96.08%

Reading time analysis. We fit a linear mixed-effects model to the first pass reading time of the second word of correctly answered critical 'yes' trials using the lme4 package in R (Bates et al., 2015; R Core Team, 2015). To be included in the analysis, the second word must have been fixated upon; six trials were discarded due to the participant not fixating on the second word (0.09%). Trials without a fixation on the first word were included, which accounted for 392 trials (5.70%). Skipping the first word did not vary by language condition or congruency, but grammatical trials were more likely to be skipped than ungrammatical trials, with grammatical trials being 3 times more likely to be skipped than ungrammatical trials. As in the accuracy analysis, the fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, English and Spanish proficiency scores as obtained on the lexTALE and the lexTALE-ESP tests respectively, and our main variables of interest: language of the second word, language congruency, and grammaticality, as well as all interactions of those three variables. All variables were centered before being added to the model.

Random effects were taken into account for both subject and item. Subject referred to the participant; item referred to the second word. Each participant was exposed to an item at most once in the experiment (participants were not exposed to all items, due to each word pair appearing in one of four possible language conditions, of which a participant was exposed to one). As in the accuracy analysis, we followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for word length of the second word, word frequency of the second word, word length of the first word, language of the second word, grammaticality, and trial number; for items, a random intercept, and slopes for English proficiency, language congruency, and trial number (Figure 7).

```
Model <- lmer(FPRT ~ TrialNumber  +
                  EnglishProficiency +
                  SpanishProficiency +
                  WordLength +
                  LogWordFrequency +
                  LogSpilloverFrequency +
                  SpilloverLength +
                  LanguageCongruence*Grammaticality*WordLanguage +
                  (1 |subject) +
                    (0 + WordLength |subject) +
                    (0 + LogWordFrequency |subject) +
                    (0 + SpilloverLength |subject) +
                    (0 + Grammaticality |subject) +
                    (0 + WordLanguage |subject) +
                    (0 + TrialNumber |subject) +
                  (1 |item) +
                    (0 + EnglishProficiency |item) +
                    (0 + LanguageCongruence |item) +
                    (0 + TrialNumber |item),
                  data=bilingual, REML=F)
```

Figure 7: Linear mixed-effects model for reading time analysis in Experiment 1.

Of our main variables of interest, we found a grammaticality effect, $\beta = 12.536$, $t(6848) = 2.087$, $p = 0.04$, showing that nouns were read faster in a grammatical word pair as compared to an ungrammatical word pair. No significant interactions were found between grammaticality and other variables of interest. Several interactions, though not significant, had large beta values; grammaticality and language, $\beta = 10.485$, $t(6848) = 1.106$, $p = 0.27$, and grammaticality, congruency, and language, $\beta = 14.280$, $t(6848) = 0.957$, $p = 0.34$, suggesting that some differences may exist across language conditions. Specifically, the interaction between language and grammaticality suggests a stronger grammaticality effect when the second word is in Spanish as compared to English. Similarly, the three-way interaction between grammaticality, congruency, and language suggests that the strongest grammaticality effect is seen in language-switch trials when the second word is in Spanish.

Most relevant to our question of interest, though the interaction between grammaticality and congruency was not significant, $\beta = 7.832$, $t(6848) = 1.068$, $p = 0.29$, its large beta value suggests there may be differences in the grammatical predictability in same language and mixed language trials. To see whether grammatical predictability is truly language-independent, we must compute the size of its effect in trials with and without a language switch, and see how they compare, which we we did in the next analysis.

In regards to language, we found a main effect, $\beta = 39.989$, $t(6848) = 3.863$, $p < .001$. As expected, Spanish words were read more slowly than English words, due to the fact that most of our participants grew up in the United States speaking Spanish only at home. We found a significant language congruency effect, $\beta = 19.574$, $t(6848) = 5.260$, $p < .001$, sometimes referred to as a language switch cost, referring to the fact that words are read slower following a language switch than when following a word of the same language. Figure 8 shows the 95% confidence interval for the main variables of interest, in milliseconds; the dotted line indicates the baseline (English, congruent, grammatical).

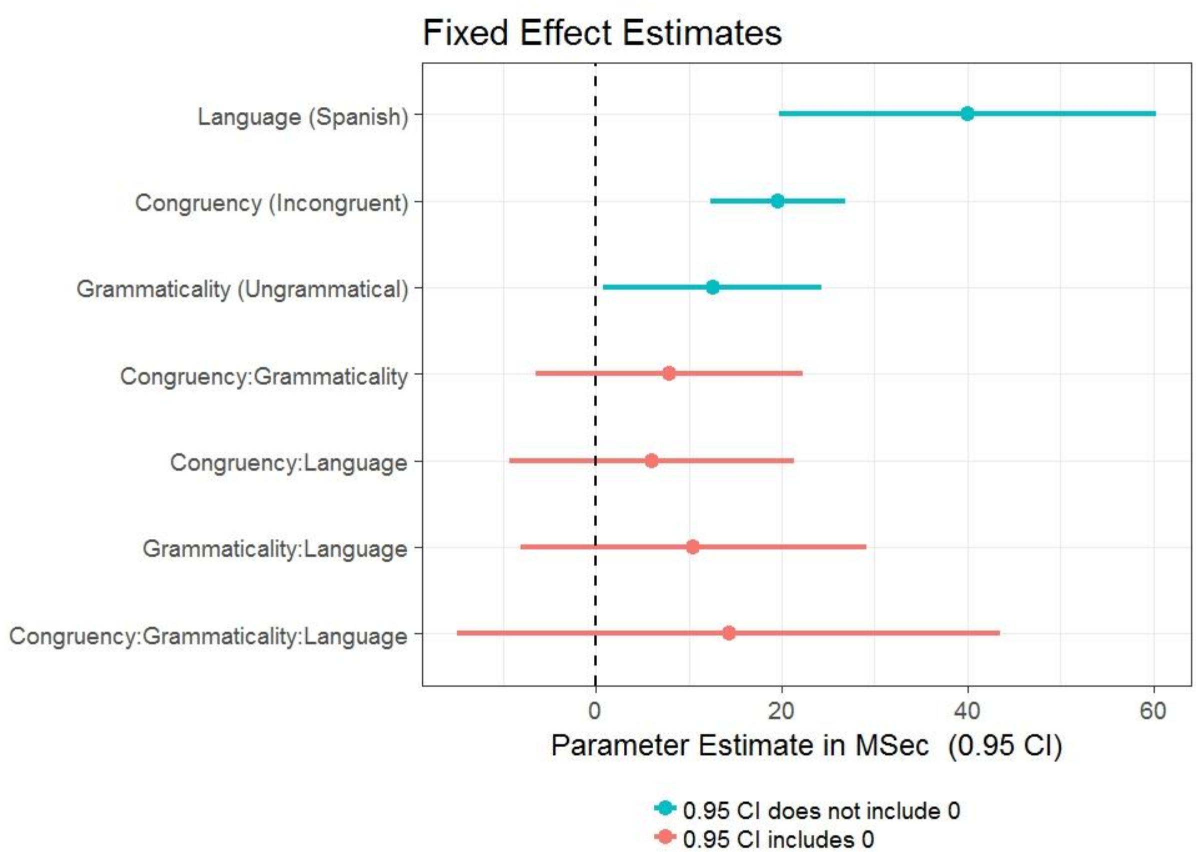


Figure 8: Fixed effect estimates for reading time in Experiment 1.

Of the control variables, we found main effects for the frequency and length of the second word, with more frequent and shorter words being read faster. (Frequency of Word 2: β = -26.299, $t(6848)$ = -5.909, $p < .001$; Length of Word 2: β = 56.205, $t(6848)$ = 12.539, $p < .001$) We did not find spillover effects nor any practice effect. We found a significant Spanish proficiency effect, β = -28.965, $t(6848)$ = -2.311, p = 0.03, showing that as Spanish proficiency increased, reading speed decreased. English proficiency failed to reach significance.

Table 5 shows the results for the fixed variables of the model shown in Figure 7. *P*-values were generated using the lmerTest package (Kuznetsova et al., 2016).

Table 5: Results of the linear mixed-effects model for reading time in Experiment 1. Significant results are bolded.

Variable	β	*t*-value	*p*-value
Intercept	**394.150**	**32.135**	**<.001**
TrialNumber	-4.945	-0.585	0.56
WordFrequency	**-26.299**	**-5.909**	**<.001**
WordLength	**56.205**	**12.539**	**<.001**
SpilloverFrequency	4.945	1.618	0.11
SpilloverLength	1.606	0.416	0.68
EnglishProficiency	-18.669	-1.490	0.14
SpanishProficiency	**-28.965**	**-2.311**	**0.03**
Language [Spanish]	**39.989**	**3.863**	**<.001**
Congruency [Incongruent]	**19.574**	**5.260**	**<.001**
Grammaticality [Ungrammatical]	**12.536**	**2.087**	**0.04**
Congruency*Language	5.963	0.762	0.45

Congruency*Grammaticality	7.832	1.068	0.29
Grammaticality*Language	10.485	1.106	0.27
Congruency*Grammaticality*Language	14.280	0.957	0.34

In Figure 9, we show the values of the predicted reading times as calculated using the coefficients shown in Table 5 for our three main variables of interest, language, language congruence, and grammaticality. The effect of grammaticality can be seen for all language combinations, with words in grammatical contexts read faster than words in ungrammatical contexts. Words following a language change are read slower than words not following a language change (congruency). This can be seen by comparing the English-English and Spanish-English language pairs, as well as the Spanish-Spanish and English-Spanish language pairs—the trend was seen for both English and Spanish words. Though the predicted times appear to show variability in the effect of grammaticality, interactions were not significant.

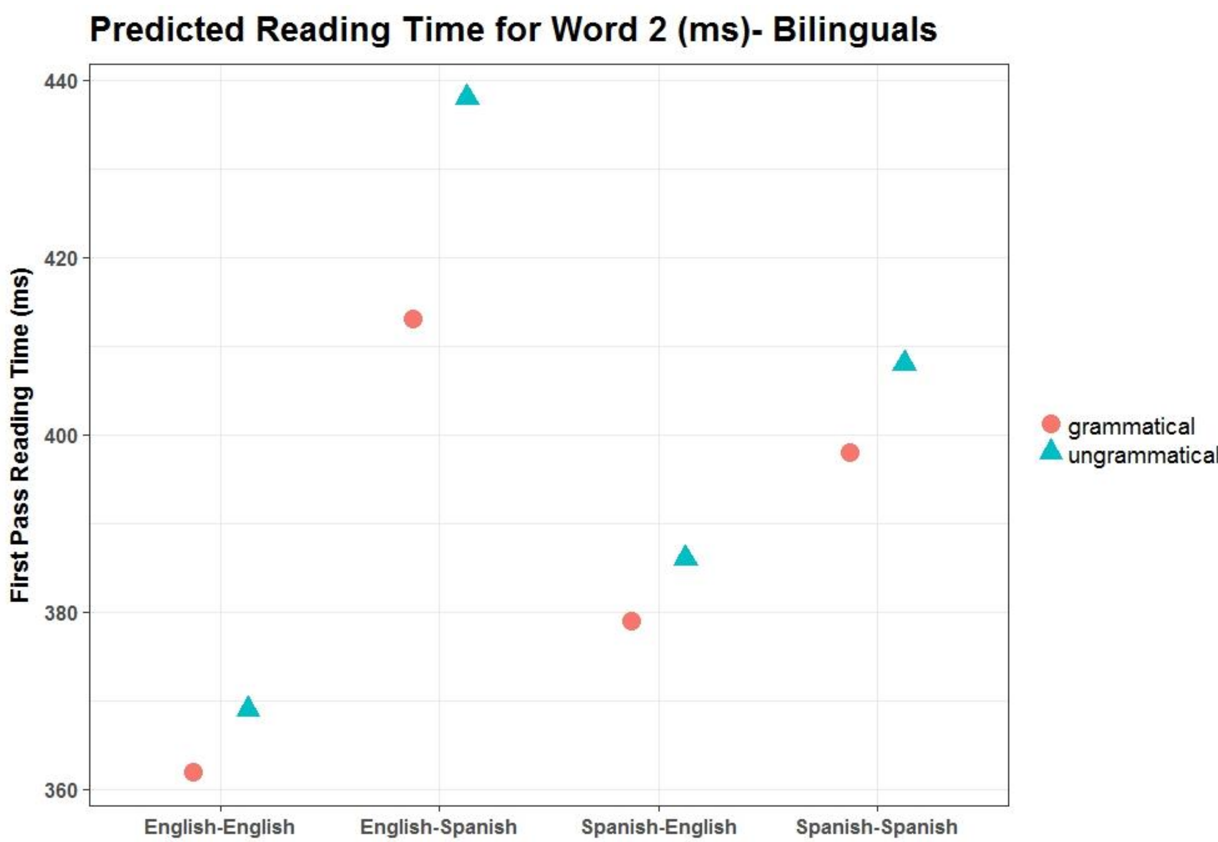


Figure 9: Point estimates for reading times given by linear mixed-effects model in Experiment 1.

Probability of an eye movement regression analysis. 9.47% of accurate trials included a regression from the second word back to the first. To see if these regressions affected the results of the reading time analysis, we looked at the probability that a trial included an eye movement regression back to the first string of characters. Table 6 shows the percent of trials with an eye movement regression as separated by language, language congruency, and grammaticality.

Table 6: Probability of an eye movement regression in Experiment 1.

	Exp 1: Bilinguals	
English-English	8.13%	
	det-N	9.74%
	adv-N	6.51%
Spanish-English	10.19%	
	det-N	11.27%
	adv-N	9.10%
English-Spanish	11.01%	
	det-N	12.40%
	adv-N	9.57%
Spanish-Spanish	8.61%	
	det-N	10.13%
	adv-N	7.05%
Same Language	8.37%	
	det-N	9.93%
	adv-N	6.78%
Mixed Language	10.60%	
	det-N	11.84%
	adv-N	9.33%
Word 2 (English)	9.14%	
	det-N	10.49%
	adv-N	7.79%
Word 2 (Spanish)	9.80%	
	det-N	11.27%
	adv-N	8.30%
ALL	9.47%	
	det-N	10.88%
	adv-N	8.04%

One pattern which is evident for all language conditions is that eye movement regressions are more common among grammatical trials as compared to ungrammatical trials. However, one factor which may be directly causing this is the fact that determiners and adverbs do not share the same word properties. When choosing the stimuli, the most common stimuli in each category were chosen—this leads to critical trials that differ in regards to first word frequency and length, both of which affect the reading time of a word. (Determiners are both on average shorter than adverbs, and also more frequent.)

Since determiners have a shorter length than determiners (average length of determiners: 5.2 characters; average length of adverbs: 9.0 characters), they take less time to read. Participants are rewarded for being fast during the experiment, so once the first string is read, they might go on to the next while still deciding whether or not it is a word, due to the shorter length. Once they determine that the second string is a word, they return back to the first string to make the decision without relying solely on memory. The fact that participants are more likely to make a regression when the first word is short instead of medium length supports this idea; participants are also more likely to make a regression when the word is long (Figure 10). The higher rate of eye movement regressions at longer words may be due to lowering frequency of the first word as the word length increases (Figure 11).

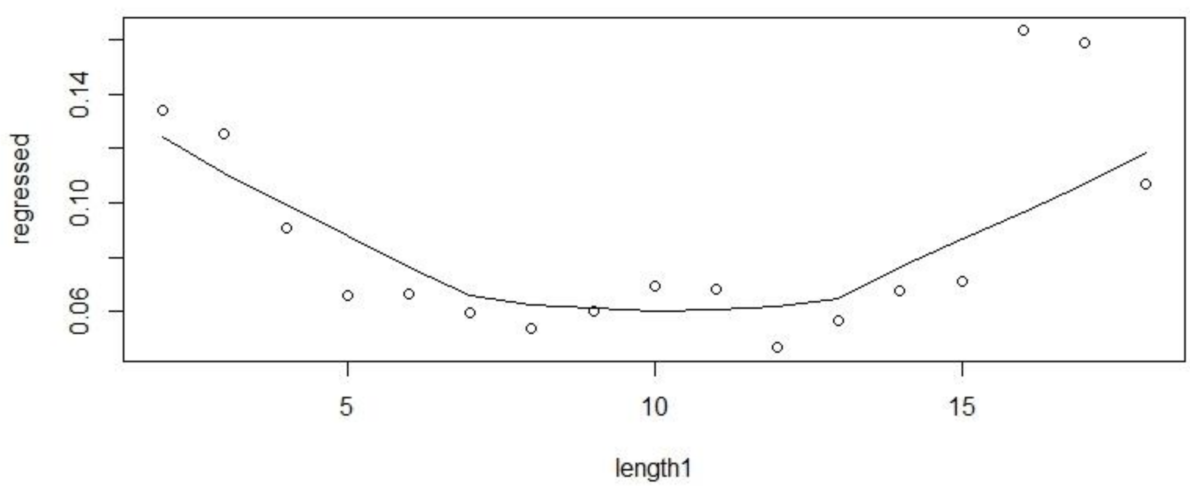


Figure 10: Probability of an eye movement regression based on length of first word in Experiment 1.

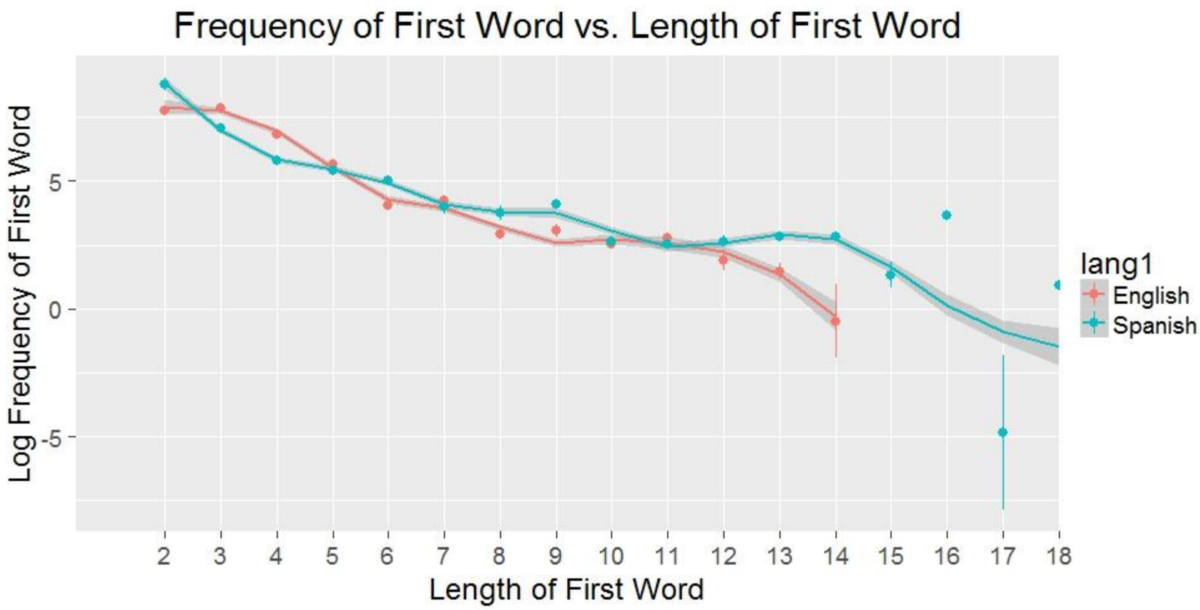


Figure 11: Frequency drops as length increases for stimuli in Experiment 1.

To analyze the data, we performed a logistic mixed-effects regression using whether or not a participant made an eye movement regression as our dependent variable. Trials with a regression were considered hits, and trials without a regression were considered misses. As in the previous analyses, the fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, English and Spanish proficiency scores as obtained on the lexTALE and the lexTALE-ESP tests respectively, and our main variables of interest: language of the second word, language congruency, and grammaticality, as well as all interactions of those three variables. All variables were centered before being added to the model.

Random effects were taken into account for both item and subject. We again followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for word length of the first word, frequency of the first word, language of the second word, language congruency, and trial number; for items, a random intercept, and slopes for language congruency and trial number (Figure 12).

```
Model <- glmer(EyeRegression ~ TrialNumber +
                      EnglishProficiency +
                      SpanishProficiency +
                      WordFrequency +
                      WordLength +
                      SpilloverFrequency +
                      SpilloverLength +
                      WordLanguage*LanguageCongruence*Grammaticality +
                      (1 |subject) +
                        (0 + SpilloverFrequency |subject) +
                        (0 + SpilloverLength |subject) +
                        (0 + WordLanguage |subject) +
                        (0 + LanguageCongruence |subject) +
                        (0 + TrialNumber |subject) +
                      (1 |item) +
                        (0 + LanguageCongruence |item)+
                        (0 + TrialNumber |item) +
                    data=bilingual, family= binomial(link = 'logit'),
                    control=glmerControl(optimizer = 'bobyqa',
                    optCtrl=list(maxfun=100000)))
```

Figure 12: Logistic mixed-effects model for probability of an eye movement regression in Experiment 1.

Mixed language trials were more likely to have an eye movement regression than same language trials, β = 0.323, $t(6850)$ = 3.223, p = 0.001, showing a language congruency effect. A main effect of trial number was found, with participants being more likely to make a regression early in the experiment as compared to later in the experiment, β = -0.688, $t(6850)$ = -5.144, p <.001. A main effect of length of Word 1 was found, with shorter Word 1 words being more likely to be regressed to, β = -0.262, $t(6850)$ = -2.852, p = 0.004). As mentioned before, when looking at the pattern of regressions based only on word length, there was also a tendency for longer Word 1 words to be followed by a regression; however this pattern is not linear (Figure 10), and there are more short words with eye movement regressions following than longer ones, and so the model predicts that shorter Word 1 lengths will lead to eye movement regressions. Finally, longer second words were more likely to be regressed from than shorter second words, β = 0.135, $t(6850)$ = 2.420, p = 0.02.

Spanish and English words were equally probable to be followed by eye movement regressions, so no main effect of language was found. No main effect of grammaticality was found, despite the fact that the numeric pattern is seen across all language combinations (Table 6). This is likely due to the fact that length of Word 1 was significant, and that grammaticality is correlated with length of Word 1, $r(1149) = .55$, $p < .001$. No other main effects or interactions were found.

Reading time analysis for trials without an eye movement regression. No main effect of grammaticality was found in the probability of eye movement regression logistic regression. However, we cannot definitively state that grammaticality is not seen in eye movement regressions, due to the fact that a main effect of Word 1 length was found, and Word 1 length and grammaticality are not independent. It is not clear if regressions would affect first pass reading time. They might not, as it is impossible to know when exactly participants are reading as opposed to deciding whether or not a string is a word. Though grammaticality and spillover length remain variables in the linear regression, by removing trials with eye movement regressions we remove noise due to decision to regress, which is strongly linked to the length of the first word (Figure 10). To see if eye movement regressions affected the results of the reading time analysis, we ran a linear mixed-effects regression using first pass reading time as the dependent variable only on trials without a regression.

As in the previous analyses, the fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, English and Spanish proficiency scores as obtained on the lexTALE and the lexTALE-ESP tests respectively, and our main variables of interest: language of the second word, language congruency, and grammaticality, as well as all interactions of those three variables. All variables were centered before being added to the model.

Random effects were taken into account for both item and subject. We again followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for word length of the second word, word frequency of the second word, language of the second word, grammaticality, and trial number; for items, a random intercept, and slopes for English proficiency, language congruency, and trial number (Figure 13).

```
Model <- lmer(FPRT ~ TrialNumber +
                     EnglishProficiency +
                     SpanishProficiency +
                     WordFrequency +
                     WordLength +
                     SpilloverFrequency +
                     SpilloverLength +
                     WordLanguage*LanguageCongruency*Grammaticality +
                     (1 |subject) +
                       (0 + WordFrequency |subject) +
                       (0 + WordLength |subject) +
                       (0 + WordLanguage |subject) +
                       (0 + Grammaticality |subject) +
                       (0 + TrialNumber |subject) +
                     (1 |item) +
                       (0 + EnglishProficiency |item) +
                       (0 + LanguageCongruency |item) +
                       (0 + TrialNumber |item),
                   data=bilingual[no regressions], REML=F)
```

Figure 13: Linear mixed-effects model for reading time analysis excluding trials with an eye movement regression in Experiment 1.

As stated, the main reason to look at this model was to see if the main effect of grammaticality would persist if the trials without regressions were included, as grammaticality may have been an artefact due to differences in reading time due to the word length of the first word. The model produced by looking at trials without eye movement regressions was quite similar to the one that included trials with eye movement regressions, with similar effects among our variables of interest, but there were a few differences in some of the control variables.

As before, a main effect of grammatical predictability was found, β = 11.614, $t(6197) = 1.911$, p = 0.05, with nouns being read faster following a determiner as compared to an adverb. Mixed language trials were read more slowly than same language trials, β = 22.079, $t(6197) = 6.106$, $p < .001$, showing a language switch cost. Trials with the second word in Spanish were read more slowly than those with the second word in English, β = 39.836, $t(6197) = 3.688$, $p < .001$. No interactions among the variables of interest were found, though large beta values were found for the interactions between language and grammaticality, β = 61.717, t(6197) = 12.547, p < .001, and language, grammaticality, and congruency, β = 61.717, t(6197) = 12.547, p < .001, suggesting that grammatical predictability effects may be different in same and different language trials.

In regards to the variables controlling for extra factors, similar effects were found regarding word length, word frequency, and Spanish proficiency. Reading time increased when the length of the second word increased, β = 61.717, t(6197) = 12.547, p < .001, when the frequency of the second word decreased β = -27.795, t(6197) = -6.008, p < .001, and when Spanish proficiency was lower, β = -30.772, t(6197) = -2.247, p = 0.03.

Unlike the model including trials with eye movement regressions, a main effect of trial number was found, with participants getting faster as the experiment progressed, β = -17.007, t(6197) = -2.733, p = 0.008. A main effect of length of the first word was also found, with longer Word 1 words being associated with shorter reading times of Word 2, β = -6.752, t(6197) = -1.966, p = 0.05. No spillover effects were found. Table 7 shows the results for the fixed variables of the model shown in Figure 13.

Table 7: Results of the linear mixed-effects model excluding trials with an eye movement regression in Experiment 1. Significant results are bolded.

Variable	β	*t*-value	*p*-value
Intercept	**405.550**	**30.415**	**<.001**
TrialNumber	**-17.007**	**-2.733**	**0.008**
WordFrequency	**-27.795**	**-6.008**	**<.001**
WordLength	**61.717**	**12.547**	**<.001**
SpilloverFrequency	4.552	1.528	0.13
SpilloverLength	-4.862	-1.498	0.13

EnglishProficiency	-21.276	-1.575	0.12
SpanishProficiency	**-30.772**	**-2.247**	**0.03**
Language [Spanish]	**39.836**	**3.688**	**<.001**
Congruency [Incongruent]	**22.079**	**6.106**	**<.001**
Grammaticality [Ungrammatical]	**11.614**	**1.911**	**0.05**
Congruency*Grammaticality	4.291	0.605	0.55
Congruency*Language	2.271	0.298	0.77
Grammaticality*Language	13.438	1.413	0.16
Congruency*Grammaticality*Language	11.818	0.817	0.41

Reading time analysis comparing grammaticality in same and mixed language trials. In the previous reading time analyses, we found a grammatical predictability effect, and no significant interaction between grammaticality and congruency. However, to definitively state that grammatical predictability is language-independent, we must compare the effect sizes for both same language and mixed language trials. To do so, we separated grammaticality into two variables instead of looking for an interaction between grammaticality and congruency; grammaticality in congruent trials, and grammaticality in congruent trials.

We fit a linear mixed-effects model to the first pass reading time of the second word of correctly answered critical 'yes' trials using the lme4 package in R (Bates et al., 2015; R Core Team, 2015). As in the reading time analysis, the fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, English and Spanish proficiency scores as obtained on the lexTALE and the lexTALE-ESP tests respectively, and our main variables of interest: language of the second word, language congruency, grammaticality in congruent trials, and grammaticality in incongruent trials. Interaction between language and congruency, language and grammaticality in congruent trials, and grammaticality in incongruent trials were also included. All variables were centered before being added to the model.

Random effects were taken into account for both subject and item. Subject referred to the participant; item referred to the second word. Each participant was exposed to an item at most once in the experiment (participants were not exposed to all items, due to each word pair appearing in one of four possible language conditions, of which a participant was exposed to one). As in the previous reading time analyses, we followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for word length of the second word, word frequency of the second word, word length of the first word, language of the second word, grammaticality in incongruent trials, and trial number; for items, a random intercept, and slopes for English proficiency, language congruency, grammaticality in incongruent trials, and trial number (Figure 14).

```
Model <- lmer(FPRT ~ TrialNumber  +
                  EnglishProficiency +
                  SpanishProficiency +
                  WordLength +
                  LogWordFrequency +
                  LogSpilloverFrequency +
                  SpilloverLength +
                  CongruentGrammaticality*WordLanguage +
                  IncongruentGrammaticality*WordLanguage +
                  LanguageCongruence*WordLanguage +
                  (1 |subject) +
                    (0 + WordLength |subject) +
                    (0 + LogWordFrequency |subject) +
                    (0 + SpilloverLength |subject) +
                    (0 + IncongruentGrammaticality |subject) +
                    (0 + WordLanguage |subject) +
                    (0 + TrialNumber |subject) +
                  (1 |item) +
                    (0 + EnglishProficiency |item) +
                    (0 + LanguageCongruence |item) +
                    (0 + IncongruentGrammaticality |item) +
                    (0 + TrialNumber |item),
                  data=bilingual, REML=F)
```

Figure 14: Linear mixed-effects model for reading time analysis comparing grammaticality in same and mixed language trials in Experiment 1.

Of our main variables of interest, we found a grammaticality effect in incongruent trials, *β* = 16.564, *t*(6847) = 2.070, *p* = 0.04, showing that nouns were read faster in a grammatical mixed language word pair as compared to an ungrammatical mixed language word pair. Grammaticality in congruent trials did not reach significance, *β* = 8.644, *t*(6847) = 1.303, *p* = 0.19, suggesting that nouns are read equally fast following determiners and adverbs in same language word pairs. Though grammaticality did not reach significance in same language trials, the grammaticality effect in mixed language trials suggests that grammatical predictability is language-independent.

Language did not interact with grammaticality in congruent trials, and the interaction was not significant with grammaticality in incongruent trials. However, the large beta value for the interaction between language and grammaticality in incongruent trials suggests that stronger grammatical predictability effects may be present when the second word was in Spanish as compared to when it was in English (*β* = 16.659, *t*(6847) = 1.339, *p* = 0.18).

As in the previous reading time model, we found a language effect, $\beta = 39.897$, $t(6847) = 3.868$, $p < .001$, with Spanish words being read more slowly than English words. There was also a significant language congruency effect, $\beta = 19.617$, $t(6847) = 5.292$, $p < .001$, showing that nouns following a language switch were read slower than nouns following a word of the same language. No interaction was found between language and congruency. Figure 15 shows the 95% confidence interval for the main variables of interest, in milliseconds; the dotted line indicates the baseline (English, congruent, grammatical).

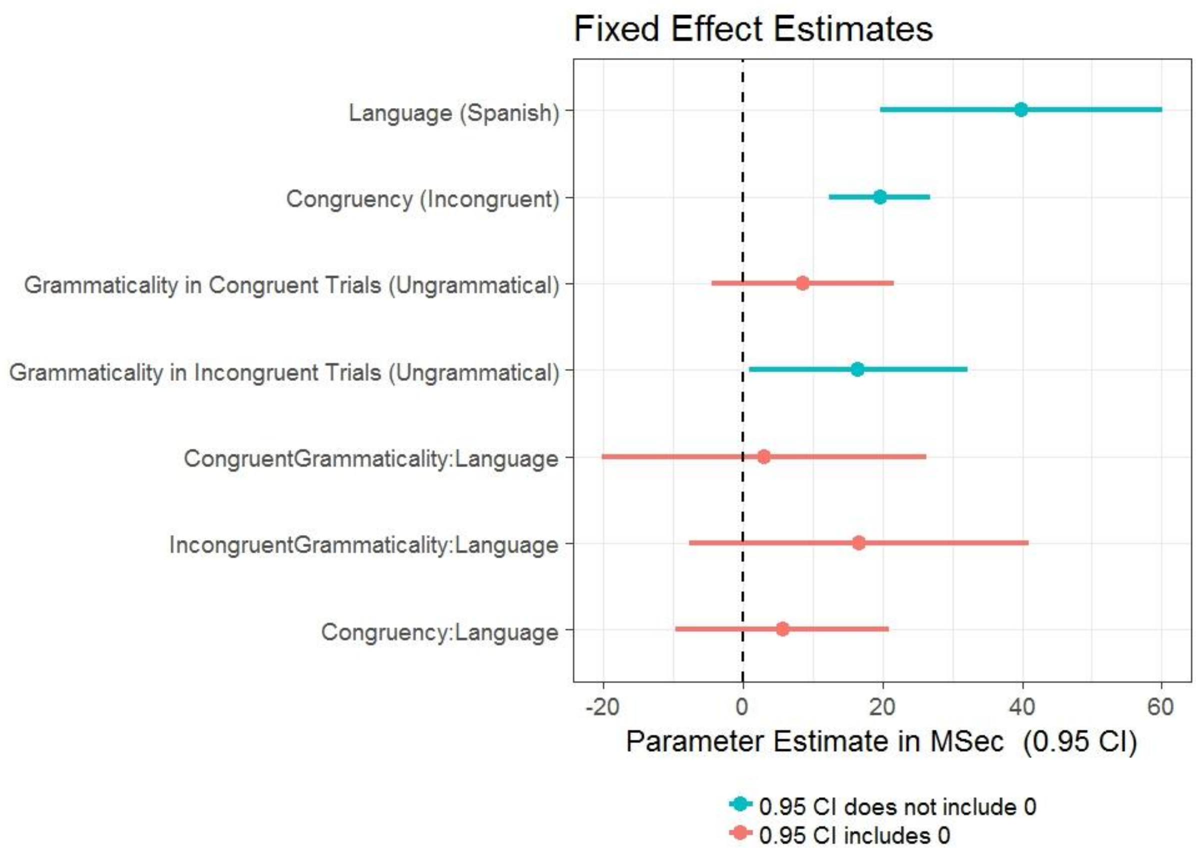


Figure 15: Fixed effect estimates for reading time analysis comparing grammaticality in same and mixed language trials in Experiment 1.

We also found main effects for the frequency and length of the second word, with more frequent and shorter words being read faster. (Frequency of Word 2: $\beta = -26.257$, $t(6847) = -5.907$, $p < .001$; Length of Word 2: $\beta = 56.260$, $t(6847) = 12.493$, $p < .001$). We did not find spillover effects nor any practice effect. There was a significant Spanish proficiency effect, β = -28.970, $t(6847) = -2.310$, $p = 0.03$, showing that as Spanish proficiency increased, reading speed decreased. English proficiency was not significant.

Table 8 shows the results for the fixed variables of the model shown in Figure 14. *P*-values were generated using the lmerTest package (Kuznetsova et al., 2016).

Table 8: Results of the linear mixed-effects model for reading time analysis comparing grammaticality in same and mixed language trials in Experiment 1. Significant results are bolded.

Variable	β	*t*-value	*p*-value
Intercept	**394.161**	**32.127**	**<.001**
TrialNumber	-4.097	-0.584	0.56
WordFrequency	**-26.257**	**-5.907**	**<.001**
WordLength	**56.260**	**12.493**	**<.001**
SpilloverFrequency	4.961	1.621	0.11
SpilloverLength	1.618	0.412	0.68
EnglishProficiency	-18.667	-1.489	0.14
SpanishProficiency	**-28.970**	**-2.310**	**0.03**
Language [Spanish]	**39.897**	**3.868**	**<.001**
Congruency [Incongruent]	**19.617**	**5.292**	**<.001**
CongruentGrammaticality [Ungrammatical]	8.644	1.303	0.19
IncongruentGrammaticality [Ungrammatical]	**16.564**	**2.070**	**0.04**
CongruentGrammaticality*Language	3.082	0.259	0.80
IncongruentGrammaticality*Language	16.659	1.339	0.18
Congruency*Language	5.725	0.734	0.46

Discussion

Lexical Decision Task

Experiment 1 looked at whether bilinguals showed predictability effects due to grammaticality in same language and mixed language word pairs. Otherwise stated, previous research had shown that monolinguals show predictability effects due to grammaticality (Staub & Clifton, 2006; Vigliocco et al., 2008); we wished to extend this research into bilinguals to see whether such effects were language-specific or language-independent. To do so, Spanish-English bilingual participants completed a two string lexical decision task while being eye-tracked, where two word trials could occur in English-English, Spanish-Spanish, English-Spanish, and Spanish-English. In particular, we looked at differences in how participants answered grammatical critical trials (det-N) as compared to ungrammatical critical trials (adv-N) and differences across different language pairs. To see whether there was evidence of grammatical predictability, we looked at the reading time of the noun, accuracy (whether or not participants answered the lexical decision task correctly), and the probability of eye movement regressions.

As stated, previous research had shown that monolinguals showed grammaticality effects, so we expected to find a main effect of predictability. As predicted, we found a main effect of grammaticality in the reading times, with participants reading nouns in grammatical language pairs faster than nouns in ungrammatical language pairs. Grammaticality effects were not present when looking at either the accuracy of response or the probability of an eye movement regression. In addition, there was no interaction between grammaticality and language of the noun, suggesting that participants are behaving the same regardless of the language.

Having found evidence of predictability due to grammaticality in reading times, we were interested in seeing whether grammaticality effects were language-specific or language-independent.

Specifically, we looked at whether there was an interaction between the grammaticality of the language pair and the congruency of the language pair (whether or not the two words of the word pair were in the same language). Finding an interaction would suggest that mixed language pairs were behaving differently to same language pairs, while not finding an interaction would suggest that grammaticality effects persisted across a language switch. We found no evidence of an interaction between grammaticality and congruency, suggesting that predictability effects were occurring across language pairs (language independent).

In addition, we compared the magnitude of the grammaticality effect in same and mixed language word pairs. Finding a smaller grammatical predictability effect in mixed language trials as compared to same language trials would suggest that though grammatical predictability was being found across languages, it may be to a lesser extent. Interestingly, we found a significant grammaticality effect in mixed language trials but not in same language trials (though the pattern of faster reading times for nouns following determiners as compared to adverbs was present in both kinds of trials). This suggests that the grammaticality effect found across all trials is being driven primarily by the mixed language trials. A greater grammatical predictability effect in mixed language trials may suggest that the effect becomes more pronounced following a language switch due to increased processing time, though this is speculative. Regardless, having found a grammatical predictability effect in mixed language trials provides evidence that predictability is occurring across language boundaries.

Finding grammaticality effects across languages provides support for theories of shared grammatical representation across languages, such as the shared syntax account. The shared syntax account holds that grammatical representations are connected directly to word lemmas which can occur within the structure, and that language is a property to which the lemmas connect; grammatical structures are not necessarily connected directly to language (Hartsuiker et al., 2004). This type of structure indicates that when a grammatical structure has the same form across language lemmas can be inserted regardless of language, i.e., code-switches, and that grammaticality would not be affected.

If instead participants had completely separate parsers, we would expect that grammaticality effects would not be present in mixed language word pairs. The separate representation account allows that the separate formulators and parsers may be linked in some way, since code-switching exists (de Bot, 1992). However, in that case we would expect grammaticality effects (if found) to be weaker in mixed language pairs, which we found no evidence of; instead, we found the opposite, that grammatical predictability effects were stronger in mixed language pairs. Overall, our evidence most strongly supports a theory of shared grammatical representation.

Other than grammaticality effects, we expected to find a language switch cost (Macnamara & Kushnir, 1971), with participants reading nouns in a mixed language word pair slower than those in same language pairs. As predicted, language congruency effect was found in reading times. Evidence of a language switch cost was also found in the accuracy and probability of eye movement regression analyses, with participants more accurate in same language trials as compared to mixed language trials, and more likely to make an eye movement regression in mixed language trials than same language trials.

We also thought it likely that participants would be faster at reading words in English than words in Spanish, as most were Heritage Speakers, and had been educated primarily in English. The lexTALE and lexTALE-ESP, vocabulary-based proficiency tests of English and Spanish respectively, also showed the same pattern, with participants tending to score higher in the English test than the Spanish test. Though this may indicate that participants were stronger in understanding English than Spanish, this may not be the case. Instead, it indicates that they have a larger vocabulary in English than in Spanish; it does not necessary mean that their Spanish grammatical knowledge was lower. Participants scored equally well on the Spanish and English oral fluency measures (the elicited imitation task).

As predicted, participants were faster to read words in English than in Spanish. This most likely indicates more experience in reading English than Spanish, since the vocabulary used in the study were chosen especially to have been learned in the first semester of a Spanish learning course. In other words, the stimuli was made up of frequent, common words. No language effect was found in either accuracy of response or the probability of making an eye movement regression analyses, suggesting that other than differences in reading time, participants treated stimuli in English and Spanish similarly. Participants also showed an equal language switch cost going from English to Spanish and going from Spanish to English, suggesting similar strength in the two languages (Meuter & Allport, 1999).

We also included variables in our analyses to account for factors known to affect reading time, the length and frequency of each of the words of the pair. We found the predicted length and frequency effects, with participants taking longer to read longer words than shorter words, and being faster to read more frequent words than less frequent words. We did not find reliable spillover effects; however, this is likely due to flaws in the stimuli. Nouns following adverbs (which tend to be longer and less frequent than determiners) were more frequent than nouns following determiners, which led to shorter reading times. This led to spillover length and frequency to be negatively correlated with the frequency of the second word, an issue of collinearity. When spillover effects were found, they were opposite to what was expected, with nouns following longer and less frequent words (such as adverbs) having shorter reading times, due to the frequency of the noun.

Code-switching Sentences Rating Task

One factor we were interested in looking at was seeing whether our participants were sensitive to code-switching norms in English-Spanish code-switching. Earlier research has shown that code-switches do not occur at all possible locations in a sentence, even for constructions in which the word order in English and Spanish is the same (Sankoff & Poplack, 1981). Furthermore, Toribio showed that bilinguals prefer code-switching confirming to norms, taking longer to read 'ungrammatical' code-switches in silent reading, and showing more false starts as well as corrections to 'grammatical' code-switches when asked to read them aloud. When asked, they preferred grammatical code-switches, often unable to articulate why (2001a).

To test whether bilinguals preferred grammatical code-switches to ungrammatical code-switches, participants were asked to complete a short acceptability judgment task where they were asked to rate how 'good' the sentence sounded to them. Bilinguals were found to rate grammatical code-switches more than ungrammatical ones, indicating that they were sensitive to English-Spanish code-switching constraints. This sensitivity was not correlated with self-reported past experience in code-switching; however, self-reported language usage measures have been found to vary by many measures, such as attitudes towards language (Bourhis, 1983). In other words, self-reported code-switching measure may not necessarily be reflecting actual code-switching use since participants are using different criteria to judge whether or not they code-switch.

In Chapter 4 we look closer at how code-switching sensitivity affected performance in the task.

Limitations

We found evidence of predictability effects due to grammaticality in bilinguals in the reading time task. However, there were some limitations to the study. First, the stimuli were not perfect; as stated above, nouns following adverbs tended to be more frequent than nouns following determiners, which was not discovered until after the experiment had been completed. This is due to the way the word pairs were made; to assure less bias, word pairs were assigned randomly by grammatical class, and then made certain to match for gender and mass/count properties (switching the second word when the word pair was ungrammatical). In addition, number was also controlled for, with plural markers added as necessary. This affects the frequency of nouns as a plural forms are most often less frequent than singular forms;

adverbs did not need to be matched by any property, which led to less instances of less frequent forms being used in those word pairs. Future studies will be able to take this more into consideration when crafting stimuli and control for the same words appearing in different conditions across participants.

Also, the nature of the stimuli used might also have affected results. The stimuli were chosen from words learned the first semester of a Spanish learning course to maximize performance in a wide variety of participants, including Spanish language learners of skill levels. This was necessary as most of our analyses look only at correct answers. However, this leads to a sample which is not necessarily representative of language as whole; using a more representative sample might expand possibilities in the future.

Finally, many words are ambiguous, in regards to grammatical class. For example, the past participle of many verbs in Spanish can be used as adjectives to describe an item affected by the action; in turn, many adjectives are read as nouns when preceded by a determiner. (E.g., *leido* is the past participle of *leer*, which means to read; *el libro leido* means the book which has been read (adjective), *el leido* can refer to an object has been read (noun).) Similar ambiguities exist in English; for example, verbs and nouns can have the same word form. (E.g., *swim* when read as a verb means to move through water by using limbs, fins, or other movements; as a noun, it means the act of doing so.) Word pairs were considered to be either grammatical or ungrammatical, based upon whether the words formed a grammatical constituent. If grammatical class is ambiguous, however, such distinctions are blurred.

Conclusions and Implications

To summarize, the most important finding is that grammatical predictability effects were found to persist across the language shift. This finding implies that at least some abstract of the grammatical representations created during incremental word-by-word processing are language-independent, in that they immediately affect processing of the incoming word regardless of its language. This indicates that bilinguals form language-independent grammatical category representations online at the word level. Previous research has found evidence of shared representations in lexical access (Marian et al., 2003), grammatical structures (Hartsuiker et al., 2004), and phonology (Roelofs, 2003), but our study is the first to establish that grammatical representations facilitate reading of words across intrasentential code-switch boundaries.

Our results most support a model of shared grammatical representations, such as the shared syntax model (Hartsuiker & Pickering, 2008). Such a model posits that grammatical representations overlap as much as possible, and we found language-independent predictability effects in word pairs that form a grammatical constituent in both English and Spanish. An additional implication is that grammatical predictability is due to abstract grammatical forms rather than previous experience.

Within a language, it can be difficult to separate the effects of abstract category-based grammaticality from past experience. For example, combinations of words forming noun phrases are both very frequent and grammatical (e.g., 'the lake' is much more frequent than 'quickly lake'), and it is difficult to separate the effect of grammaticality from token-specific past experience. But across languages, the frequency of specific cross-language determiner-noun pairs is small (perhaps zero for many pairs), and so it is striking that the magnitude of the grammaticality-like effect we found both within and across languages is similar. Such effects demand the assumption of predictions based on abstract category in predictive models.

Our findings lead to a series of new questions looking at the nature of grammatical predictability, some of which we address in this dissertation, though many still remain. In the following text, we focus on the questions that led to Experiments 2 and 3; further questions are discussed in Chapter 6.

One question is how grammatical predictability interacts with language proficiency. Our participants were Spanish-English bilingual, meaning anyone who self-identified as having learned Spanish before age 6 and was able to read a newspaper in both languages. This led to a heterogynous sample: participants ranged from children of immigrants, who spoke Spanish at home but English in their everyday lives to adults who learned English in their 20s after completing their schooling in Spanish. (Most participants were children of Spanish-speaking immigrants.) Proficiency scores varied on both languages across participants. Such variability makes it difficult to look at how proficiency affects performance since we have two variables interacting. In Experiment 2, we decided to examine the effects of proficiency more closely by looking at Spanish L2 language learners with the same methodology. This makes for a sample with little variation on English proficiency (all participants are native English speakers who have been educated in English), but with relatively high Spanish proficiency variation.

In Experiment 3, we focused on the issue of ambiguity in grammaticality. As mentioned previously, the fact that the same word may have multiple meanings, some of which transcend grammatical class leads to word pairs that can be read as either grammatical or ungrammatical. Instead of treating grammaticality as binary, we used conditional probabilities to calculate variables indicating the likelihood of the words of the word pair occurring together. This also allowed us to examine whether grammatical predictability effects are due to the reader predicting the grammatical class of the following word after reading the first word, or if instead they are due to differences in processing cost between grammatical and ungrammatical structures.

Chapter 3

Experiment 2: Spanish Language Learners

Introduction

In Experiment 1 we found evidence that grammatical predictability was language-independent in Spanish-English bilinguals. This provides evidence for theories of shared representation, such as the shared syntax account, and opens the door for new avenues of investigation. In this experiment, we wish to test the effects of linguistic proficiency on grammatical predictability.

In Experiment 1, we strove to find participants with similar linguistic knowledge in both their languages. Looking at their proficiency levels in each language shows that we were relatively successful in doing so; overall proficiency was high in both English and Spanish, though participants tended to be more proficient in English (mean: 90.25%, sd: 8.91%) than in Spanish (mean: 79.92%, sd: 10.98%). This led us to predict equal levels of grammatical predictability regardless of the language condition in the case of language-independent predictability, which was what we ultimately found.

Looking only at participants with similar proficiency levels in both their languages leaves open the question of what effect do differences in linguistic proficiency have on grammatical predictability. Predictability occurs based on previous linguistic knowledge; therefore, people with differing levels of linguistic knowledge in their two languages are likely to show different levels of grammatical predictability for each language. In this experiment, we are interested in studying participants who have different levels of proficiency in their two languages; otherwise stated, they are experts in one language (L1), but have a lower L2 proficiency level.

In Experiment 1, we sought to find participants with similar levels of proficiency in each language; most participants spoke both English and Spanish since before age 5, making them likely to show expert-like proficiency in each of their two languages. No two bilinguals are identical, and participants rarely scored exactly the same in the lexTALE and lexTALE-ESP tests. Overall, bilingual participants had similar proficiencies in their languages, averaging a difference of ~10% across languages. Individual proficiency in each language was not expected to modulate predictability effects, so we added proficiency test scores (lexTALE and lexTALE-ESP) to account for individual differences.

In this experiment, we wish to look at participants who have lower proficiency level in one language than another language. To do so, we looked at advanced English-Spanish language learners, native speakers of one language (English), and proficient in the other (Spanish), which they learned after puberty. As with bilinguals, we are not predicting Spanish (or English) proficiency to modulate grammaticality effects, as we selected participants who have similar English proficiency levels to each other (native speakers) and similar Spanish proficiency levels to each other (Spanish majors/minors taking college courses taught in Spanish). As we did in Experiment 1, we will include the lexTALE and lexTALE-ESP scores to the model to account for individual differences.

We are interested in seeing the effect of a participant having a higher proficiency in one language as compared to their other language; specifically, the difference between a language learner's native and second language. We chose to focus only on advanced learners as we do not know what level of linguistic proficiency in a language is necessary for a participant to show grammatical predictability, and advanced learners are more likely to show evidence of grammatical predictability than less skilled participants (who would have a greater difference in their proficiency levels between their languages).

To summarize, advanced English-Spanish language learners are native English speakers, and so should be at ceiling for English proficiency. In contrast, their Spanish proficiency scores are significantly lower, as they are not native speakers, though they studied Spanish for an extended period of time. If grammatical predictability is moderated by linguistic proficiency in a language, we would see differing amount of predictability effects based on the language condition, as all participants are stronger in English than in Spanish.

To our knowledge, this is the first study that directly looks at whether language learners show different processing times for whether words are of the expected lexical category both across a language switch and within the L2. Language learners are expected to show faster processing times within their L1 for words in grammatical constituents, but whether or not they show that same pattern in the L2 and across language switches depends on whether the knowledge of the languages is accessed simultaneously.

For this study, we looked at advanced English-Spanish language learners who using the same stimuli and methodology as Experiment 1. We used eye-tracking to measure how advanced Spanish L2 speakers processed grammatical and ungrammatical word pairs. Participants performed a lexical decision task where they had to judge whether or not both letter strings presented were words, regardless of whether they were in English or Spanish. Additionally, participants completed a battery of proficiency tests, as well as background questionnaires focusing on their use of code-switching.

Most importantly, we wished to see if a grammaticality effect was present across language boundaries. We expected to replicate our findings from Experiment 1 in that a grammaticality effect would be found in a participant's native language; however, we were unsure as to whether or not lower Spanish proficiency modulated predictability effects in Spanish, which would affect the other three language conditions. A lack of a grammaticality effect in Spanish would suggest that language learners are not processing oncoming words as native speakers do. Furthermore, we were interested as to whether we would find grammaticality effects in the mixed language conditions; if we did not, it would suggest that grammaticality does not persist across a language switch, supporting language-specific grammaticality effects in language learners.

We also expected to find longer reading times for Spanish words, as participants have linguistic knowledge of English and much more experience in reading English. We expected to find a language switch cost (Macnamara & Kushnir, 1971), though we were unsure as to whether or not the magnitude would be the same going from the L2 to the L1 as going from the L1 to the L2 (Meuter & Allport, 1999).

Below, we briefly review the literature on language learners that motivates these predictions, and link specific predictions to the relevant theories.

When first starting to learn a language, a language learner has no linguistic knowledge of the L2. However, after studying a language, a language learner's linguistic knowledge can be similar to that of a proficient bilingual. The types of representations are not necessarily the same, however, and the linguistic cues used in grammatical processing have been shown to be different from those of native speakers (Clahsen & Felser, 2006a). For example, unlike native speakers, language learners did not show consistent phrase structure-based ambiguity resolution strategies when resolving ambiguous relative clauses in their L2. Instead, disambiguation preferences were more influenced by the linking preposition, with learners showing no preference when the preposition was 'of' but a clear preference when the preposition was 'with' (Felser et al., 2003). Language learners also appear to be less sensitive to prosodic cues than native speakers as a way of extracting semantic information (Akker & Cutler, 2003).

Whether or not a language learner can obtain native-like competence in regards to syntax in their L2 is strongly debated; results are mixed as to what the ultimate level of attainment of a second language is (Birdsong & Vanhove, 2016). Though some researchers find native-like performance by language learners (White & Genesee, 1996), others claim equal attainment does not occur (Abrahamsson & Hyltenstam, 2009; DeKeyser et al., 2010). Complexity of the grammatical structure may be linked to ultimate attainment. In other words, language learners may show native-like processing on simple grammatical structures, but not in complex structures (Clahsen & Felser, 2006b).

Language proficiency is linked to whether or not a language learner processes language similar to a native speaker; for example, advanced Spanish L2 learners (similar to our participants) are able to use gender cues in processing an incoming word, but learners with a lower Spanish L2 proficiency are not (Dussias et al., 2013). L2 proficiency has also been linked to L2 learners' ability to process coreferents in discourse, with higher L2 proficiency speakers performing more similarly to native speakers than lower L2 proficiency speakers (Grüter et al., 2014).

The fact that language learners show native-like processing on any grammatical form relates to an interesting question: do language learners have shared syntactic representations across languages? As we discussed before, bilinguals code-switch across languages, showing moments when both languages are activated (Poplack, 1980). Similarly, syntactic priming has been found across languages, supporting a model of shared syntactic representations (Hartsuiker & Pickering, 2008). In the next section, we will look at whether language learners show evidence for shared grammatical representations, looking at code-switching behavior and evidence from syntactic priming studies.

Code-switching in Language Learners

Code-switching in bilinguals is thought to occur for a large variety of reasons. Some are linguistic in nature, such as a particular word lack in one language, or the fact that a word is more 'available' in one of the languages. Others are more social in nature, such as to emphasize group membership, draw focus to a part of a sentence, express certain emotions, as well as others (Grosjean, 1982; Appel & Muysken, 1987). In general, though code-switching is most often recorded in conversations among an established bilingual community (Poplack, 1980), or in informal conversation with balanced bilinguals in laboratory settings (Poulisse & Bongaerts, 1994). It is considered a language skill among users proficient in the users of two languages.

In comparison, code-switching in language learners is often seen in classrooms, where it is a tool to teach a majority language to speakers of a minority language, often during the instruction of other topics (Zentella, 1981; Setati et al., 2002). Code-switching is also found in L2 language classes, however, in this case, language learners are not encouraged to code-switch, since use of the L1 is highly discouraged and often not allowed in L2 language classes (Levine, 2003). Most studies have not looked at more naturalistic code-switches, ones similar to the code-switches made by bilinguals, instead focusing on the social use of code-switching and language learning.

In studies that have looked at the content of code-switches in language learners, a distinct pattern emerges. In beginning language learners, code-switches are very common, and are often words meant to call attention to an edit of the utterance that the speaker is about to make next (e.g., *I mean...*), function words, or tag words at the end of sentences. However, as they advance in their study, the number of code-switches decrease, especially among function words and editing terms (Poulisse & Bongaerts, 1994). As they advance in their course of study, code-switches more similar to what is seen by bilinguals start emerging (Liebscher & Dailey-O'Cain, 2005; Moore, 2002), especially in cases of language immersion (Arnfast & Jørgensen, 2003).

In non-observational studies, results have been more mixed. In one series of studies, language learners showed equal processing times for code-switches, no matter where they occurred in the sentence, though bilinguals showed slower processing times for ill-formed code-switches (Rakowsky, 1989). In other studies, language learners have been shown to prefer code-switches similar to ones preferred by bilinguals. In other words, more advanced learners show more preference to code-switches conforming to the same linguistic constraints. This preference was shown to interact with L2 language proficiency, with more advanced language learners showing stronger preferences for 'grammatical' code-switches than beginner language learners, suggesting that sensitivity to code-switching constraints increases with L2 proficiency (Toribio, 2001b).

Syntactic Priming in Language Learners

Among bilinguals, activation of a structure in one language leads to it being activated in their other language. Evidence of activation in the other language is seen by cross-linguistic syntactic priming. Cross-linguistic syntactic priming is a phenomenon in which a speaker of more than one language, after encountering a structure in one language, is then more likely to produce that same structure in another language (than they would have been had they not encountered the structure previously). Cross-linguistic syntactic priming is also seen in comprehension, in that after having encountered a particular structure in one language, it is then more easily processed in another language. Cross-linguistic syntactic priming has been found to occur across a number of language pairs and using a wide variety of tasks, as was discussed previously in the introduction.

Though not as extensively studied, some studies have found cross-linguistic syntactic priming in language learners. English L2 learners (Swedish L1) showed equally strong cross-linguistic priming of datives using both Swedish and English primes as well as both English and Swedish targets in a sentence completion task (Kantola & van Gompel, 2011). Though that particular study did not test effects of L2 linguistic proficiency, further research has suggested that it might moderate the effects of priming.

Stronger cross-linguistic priming was found for more proficient language learners in a sentence production task (Bernolet et al., 2013), suggesting that with greater L2 proficiency comes increased activation across languages. One cross-linguistic priming study using a sentence production task using both code-switching primes and targets (code-switching kept constant L1 -> L2) also found higher levels of priming in more advanced language learners (Koostra et al., 2012). In contrast, language learners who were less proficient in the L2 showed greater effects of priming in a comprehension study that testing cross-linguistic priming from the L1 to the L2 (Hopp, 2017), suggesting that greater proficiency in the L2 leads to more inhibition of the L1.

L2 (or L3) proficiency was not found to affect a sentence completion task looking at attachment preferences. In a set of studies, native Dutch speakers were tested with Dutch primes on Dutch as well as L2 targets, and L2 primes with Dutch or L3 targets. Equal amounts of priming were found regardless of the language of the prime and target, or self-rated proficiency in each language (Hartsuiker et al., 2016).

These mixed results leave the question of how a difference in L1 nd L2 proficiency affects the shared grammatical representations across languages open. Our study seeks to further examine the effect of different L1 and L2 proficiency levels on shared grammatical representations, by seeing if participants show a stronger grammatical predictability effect in their L1.

Revised Hierarchical Model (RHM)

One influential model of bilingual language learning and transfer is the revised hierarchical model (RHM), first proposed by Kroll & Stewart (1994). In this model, language learners show changes in how L2 words are processed as their proficiency increases. Proficient L2 learners process L2 words directly, similar to bilinguals. However, less proficient learners go through a stage of relying on word association between the L1 and L2 to access concepts using the L2. L2 words are not linked to concepts directly; they are connected to their translational equivalent and use the translational equivalent to access the word concept. Learners eventually form direct links between L2 words and concepts, and the temporary dependency on word association between the L1 and L2 diminishes as proficiency increases.

Unfortunately, there is no easy method to distinguish whether a language learner is processing L2 words directly or not, only a rough guideline that states that beginner learners do not, but advanced learners do. Additionally, it is likely that changing to processing words directly is a gradual process, depending on word use. In our study, we looked at advanced language learners and words they likely had learned early on in their course of learning Spanish, and we can say it is likely that they process the language directly, however, there is no guarantee that they do.

Some evidence for a translational stage to access L2 forms comes from the fact that learners show longer naming times in translation from L1 -> L2 as compared to L2 -> L1. This same pattern is found in both low and advanced L2 learners (Kroll & Stewart, 1994). Additionally, low proficiency language learners show interference from words with a similar lexical form but different meaning (e.g., *man-hambre* [*hambre* means 'hunger' in Spanish; *hombre* means 'man']) in a translation recognition task where they have to respond whether or not the two words have the same meaning. Advanced learners do not show the same interference from similar lexical form (Talamas et al., 1999). A similar study did find that advanced learners showed interference from lexical form; however, the effect decreases as proficiency increases (Sunderman & Kroll, 2006).

Asymmetrical Costs of Code-switching

Switching from language to language has a cost; participants are often slower in naming images following a language switch (Costa & Santesteban, 2004). Neuroimaging studies show that code-switches elicited a greater N400 than same language trials (Moreno et al., 2002), and is associated with brain areas associated with cognitive control (Abutalebi et al., 2007).

One important finding regarding switching cost is that the language learners similar to those in our study (but with greater L2 experience, because they were immersed in the L2) show greater costs when switching into the L1 from the L2 as compared to switching into the L2 from the L1 (Meuter & Allport, 1999). The difference in switch cost has been shown as supporting the Inhibitory Control model of bilingual speech production (Green, 1998). Though we will not go into the details of the model, the most pertinent claim by the model is that bilinguals have words in both languages activated when producing speech; to choose a specific word, and not its cross-language competitor, the bilingual must inhibit the activation of the competitor. The amount to inhibition necessary is related to the strength of the activation, which is related to the strength of the linguistic knowledge; by this model it is easier to speak the more dominant language because it requires less inhibition than the opposite.

To switch languages, the inhibition must be overcome; the greater the amount of inhibition, the greater the cost. Therefore, it is easier for L2 speakers (or unbalanced bilinguals) to switch into the L2 from the L1 rather than vice versa. The asymmetrical language switch cost is not seen in balanced bilinguals because equal amounts of inhibition are necessary across languages.

Language Learning and our Study

The RHM is primarily a production model; however, it is relevant to our study in that it looks at how words may be accessed by language learners. If language learners have an intermediate stage in which their L1 must be accessed before they can process a word, it could mean that we would not see the same effects in mixed language pairs. We would however expect to see higher language switch cost for switches from Spanish to English than for switches from English to Spanish (Meuter & Allport, 1999). The effects of proficiency on grammatical predictability are unclear. There is some evidence that language learners can access grammatical information from L2 words directly, and that this does not interact with proficiency (Sunderman & Kroll, 2006). However, this has not been tested in how it relates in processing grammatical phrases, only in looking at whether grammaticality category affects lexical form interference in a translation judgment task.

In our study, we wish to test whether the grammatical predictions made by language learners are language-independent or language-specific, replicating Experiment 1 but looking at a different type of participant. In particular, we wish to see whether syntactic category expectations operate across language boundaries for cases in which grammaticality is shared across languages. In other words, after encountering an English determiner like 'the', does a L1-L2 English-Spanish language learner most likely expect a noun in English such as 'book' (showing language-specific syntactic representations), or do they most likely expect a noun in either English or Spanish (showing language-independent syntactic representations)—either 'book' or 'libro'. We will examine this phenomenon looking at how advanced L2 learners process two word phrases in their L1, their L2, and in code-switches (one word in each language).

We have several hypotheses for the experiment. We expect to find that language learners read nouns following determiners faster than those following adverbs, at least for some language pairs, showing an effect of grammaticality. Participants are also expected to read English nouns faster than Spanish nouns, as they are stronger in English, showing a main effect of language. We expect to find that language learners read nouns in same language pairs faster than those in mixed language pairs, showing a language switch cost. This language switch cost is likely to interact with language, with a larger language switch cost in Spanish-English than English-Spanish pairs, reflecting an asymmetrical language switch cost.

To see if different levels of proficiency in the L1 and the L2 affect predictability, we are interested in seeing if participants show equal amounts of grammatical predictability when the noun is in Spanish as they do when the noun is in English. An interaction between language and grammaticality, with participants showing a greater grammaticality effect for English nouns than Spanish nouns, would suggest that grammatical predictability is moderated by proficiency.

We are most interested in seeing whether participants show a grammatical predictability effect in mixed language pairs, directly pertaining to our main question of interest—is grammatical predictability language-specific or language-independent? If it is language-specific, we expect to find an interaction between language congruency and grammaticality; if it language-independent, there will be no interaction between language congruency and grammaticality.

Finally, we are interested to see whether we will find a three-way interaction between language, language congruency, and grammaticality, especially if we find an interaction between grammaticality and language and an interaction between grammaticality and language congruency. Such an interaction would likely suggest that grammatical predictability is only found in English-English word pairs, suggesting that language learners are showing language-specific grammatical predictability in their native language but not their second language.

Method

Participants

Fifty fluent advanced L2 learners of Spanish from the University of Michigan participated in the experiment. All participants were native speakers of English, and were either Spanish majors or minors and taking upper-level courses taught in Spanish. Participants received $15 base pay as well as bonuses based on speed and accuracy, for a total average payment of $22.75 (range: $20.12 : $24.65, sd = $1.09).

The mean age of the participants was 21.18 years (range: 18 : 23). Most participants had been studying Spanish for 6-9 years, having started in middle school and continued in college.

Materials

Word stimuli. Participants saw the same stimuli as in Experiment 1.

Language assessments and questionnaires. Most language assessments and questionnaires were the same as used in Experiment 1: lexTALE-ESP, lexTALE, ACSES, LHQ 2.0, code-switching assessment, and Spanish EIT. Unlike Experiment 1, participants were not given the English EIT, due to the fact that they were all native speakers and native speakers score near ceiling on elicited imitation tasks. Another L2 (Spanish) proficiency task was added instead, a Spanish grammar test designed to test forms which advanced language learners still often show mistakes in.

The Spanish grammar test was created by Lorenzo Amaya-Garcia and is a 45-question multiple choice test (3 choices per question), where participants are asked to fill in the blank in sentences in a paragraph written in Spanish. The test was designed to test grammatical forms which English learners of Spanish often do not use properly, because they either have no English analogue or because they are rare (e.g., subjunctive tenses, impersonal third voice, the conditional, the future tense, etc.). In particular, it is designed to differentiate between high-level language learners of Spanish who may score similarly to each other on the lexTALE-ESP. The test was given in paper format.

Procedure

Participants were asked to complete the same two word lexical decision task as in Experiment 1, with the same incentive formula (Figure 4). Though the formula to calculate number of points scored per trial was the same as in the previous experiment, participants were given a lower base fee for coming in to complete the study ($15 instead of $25). This is due to the fact that language learner participants were easier to find and recruit, as the task was the same in both experiments.

After completing the two word lexical decision task, participants completed the language assessments and surveys in the following order: lexTALE-ESP, lexTALE, Spanish grammar test, code-switching sentences assessment task, ACSES, LHQ 2.0, Spanish EIT. No feedback was given following the assessments.

Scoring

Language proficiency measures. 50 participants provided data on the lexTALE, lexTALE-ESP, code-switching sentence judgment task, ACSES, and LHQ 2.0 proficiency tasks, and it was analyzed in the same manner as the bilingual participants in Experiment 1.

Spanish EIT. The data from 3 participants' Spanish EIT are missing due to recording issues, leading those 3 participants to be discarded from all analyses for the EIT. The other 47 participants' data was used for the analyses. As for the bilinguals, each sentence was scored on a value of 0 (no information repeated) to 4 (sentence perfectly formed); mid-values show degrees of success in repeating the sentence. All subjects were rated by the same rater, and a different rater checked 30% of the participants. Ratings between the raters were shown to be correlated to a high degree $r(13) = 0.97$, $p < .001$, showing that rating was consistent.

Spanish grammar test. A question was scored as correct if a participant chose the correct answer, and incorrect if an incorrect answer was chosen or no answer was selected. The maximum possible score was 45; once a participant was given a final score, it was converted to a percentage score to make it easier to compare against the lexTALE-ESP.

Lexical decision task. For our analyses, we looked only at critical grammatical and ungrammatical trials. Filler trials and 'No' trials were not analyzed. Due to the fact that mean reading times varied by participant (range: 228ms : 603ms), outliers were calculated by participant. Words with reading times less than 50 ms or more than three standard deviations from that participant's mean reading time were dropped from analysis (2.53% of trials), leaving 7018 total trials used in the analyses.

We performed several sets of analyses. We first conducted an accuracy analysis. As with the bilingual participants in Experiment 1, we used each participant's average percentage correct (APC) by language condition x grammaticality as our dependent variable, for a total of eight values per participant (400 total). We also analyzed the first pass reading time of the second word as our primary measure. Only correct 'yes' grammatical and ungrammatical trials were analyzed, lowering the total number of trials analyzed to 6717. We also analyzed the probability of whether or not a participant made an eye movement regression from the second word back to the first word; as in the reading time analysis, only correct trials were scored. Trials that included an eye movement regression were scored as a 'hit', with a value of '1', and trials without an eye movement regression were scored as a 'miss', with a value of '0'. A final analysis looked at the first pass reading time of trials without regressions, lowering the total number of trials to 6334.

Results

Language Proficiency Measures

We looked at a total of 52 correlations and planned comparisons, so as to avoid false positives, a Bonferroni correction was applied; probability values have to be less than $9.62*10^{-4}$ to be considered significant. Table 9 shows the average score for each measure.

Table 9: Language proficiency and use means in Experiment 2.

lexTALE score	**97.25%**
English self-rating [1:7]	**6.91**
lexTALE-ESP score	**63.13%**
Spanish self-rating [1:7]	**6.25**
Spanish Grammar Test	**68.31%**
Spanish EIT [1:4]	**2.79**
Estimated Percent Time Code-switching	**21.46%**
Percent Time Code-switching with Top Conversational Partners	**17.37%**
Code-switching Positivity Rating [1:7]	**4.82**
Grammatical Code-switching Sentence Rating	**5.66**
Ungrammatical Code-switching Sentence Rating	**4.40**

ACSES and LHQ 2.0. Across all 50 participants, the mean self-rating for English was 6.91 (range: 6 : 7, sd = 0.21), and the mean self-rating for Spanish was 4.93 (range: 3 : 6.25, sd = 0.73). Participants consistently rated themselves as more proficient in English than Spanish, $t(49) = 19.04$, $p < .001$.

The mean self-rating for the percentage of time spent code-switching was 21.46% (range: 0% : 80%, sd = 18.40%). The estimated percent mean time spent code-switching score with top conversational partners was 17.37% (range: 0% : 91.67%, sd = 18.57%). The two scores were not significantly correlated, and nor were they significantly different, $t(49) = 1.23$, $p = 0.11$.

Participants were generally neutral or positive towards code-switching, giving a mean score of 4.82 (range: 1 : 6.7, sd = 1.28). Code-switch positivity scores were not correlated with percentage of time spent code-switching, nor self-ratings of Spanish or English.

lexTALE and lexTALE-ESP. The lexTALE and lexTALE-ESP were used as our measures of basic proficiency during the task. LexTALE scores were near ceiling, ranging from 85.00% to 100%, (mean: 97.25%; sd: 2.38). LexTALE-ESP scores were lower and showed greater variability, ranging from 48.33% to 78.33%, (mean: 63.13%, sd: 7.17). This is not surprising, as participants were native speakers of English, and Spanish was their L2. The test scores reflected this fact, with participants scoring significantly higher in the lexTALE (English) than the lexTALE-ESP (Spanish), $t(49) = 33.93$, $p = 1.07*10^{-35}$.

The lexTALE was positively correlated with participants' self-rating of English, $r(48) = 0.70$, $p = 1.91*10^{-8}$, indicating that participants were generally accurate in their self-rating. This correlation was driven by an outlier, a participant who scored 85% on the lexTALE and rated themself as being less proficient in English; removing that participant from the data caused the correlation to disappear. Participants were not similarly accurate with their self-rating of Spanish, as lexTALE-ESP score and Spanish self-rating score were not correlated. Neither lexTALE-ESP nor lexTALE scores were correlated with either code-switching use measure, nor with the positivity rating towards code-switching.

Code-switching sentence ratings. Across all participants, the average grammatical sentence score was 5.66 (range: 1 : 7, sd = 1.42), and the average ungrammatical sentence score was 4.40 (range: 1 : 6.6, sd = 1.35). Participants who rated grammatical sentences as higher also tended to rate ungrammatical sentences as higher, $r(48) = 0.68$, $p = 5.37*10^{-8}$. Participants rated grammatical code-switching sentences higher than ungrammatical ones, $t(49) = 8.03$, $p = 1.69*10^{-10}$, showing they accepted sentences conforming to known Spanish-English constraints as more grammatical than ones which violated those constraints. Grammatical and ungrammatical sentence judgment scores were not correlated with any proficiency measure, nor with time spent code-switching measures, or positivity towards code-switching measures.

Spanish grammar test. The average score on the grammar test was 68.31% (range: 42.22% : 95.56%, sd = 13.68%). Scores on the grammar test were positively correlated with lexTALE-ESP scores, $r(48) = 0.51$, $p = 1.34*10^{-4}$, showing that participants who scored high on one Spanish proficiency measure scored high on the other. Grammar test scores were not correlated with any other measure previously mentioned.

Spanish EIT. The Spanish EIT was the final measure of proficiency taken of Spanish. For the 47 participants for which we had data, the average score (out of a maximum of 4) was 2.79 (range: 1.00: 3.73, sd = 0.60), indicating a wide range of Spanish proficiency. The Spanish EIT score was correlated with both other Spanish proficiency measures, showing that all proficiency measures found similar results. (With the lexTALE-ESP score, $r(45) = 0.50$, $p = 3.16*10^{-4}$; with the grammar test, $r(45) = 0.60$, $p = 7.15*10^{-6}$). Spanish EIT scores were not correlated with any self-reported measure of proficiency or code-switching, nor with the scores on the code-switching sentence judgment task.

Lexical Decision Task

Accuracy analysis. Average accuracy was 95.71% across critical trials, with highest accuracy for English-English grammatical trials (99.67%), and lowest for Spanish-Spanish ungrammatical trials (91.73%). We were unable to fit a logistic regression model that accurately modeled the data, so we instead used a linear regression model with average percent correct as the dependent variable, including subject as a random variable, but not item.

As in the analyses in Experiment 1, the variables in the model included those which corresponded to our variables of interest: grammaticality, language congruency, and language of Word 2. All interactions including the variables was included. Each was a binary variable and centered.

Proficiency as measured by the lexTALE and the lexTALE-ESP tests were included as fixed variables to take into account English and Spanish proficiency, respectively. Proficiency was centered before being added to the model.

As before, we used a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subject, a random intercept, as well as slopes for the language of the second word and language congruency (Figure 16).

```
Model <- glmer(APC ~
                  EnglishProficiency +
                  SpanishProficiency +
                  WordLanguage*LanguageCongruence*Grammaticality +
                  (1 |subject) +
                    (0 + LanguageCongruence |subject) +
                    (0 + WordLanguage |subject) +
                  data=learner, REML = F)
```

Figure 16: Linear mixed-effects model for accuracy analysis in Experiment 2.

An effect of grammaticality was found, with participants more likely to be accurate when the noun was following a determiner as compared to an adverb, β = -0.024, $t(387)$ = -5.028, p < .001. Trials with the second word in English were more accurate than trials with the second word in Spanish, β = -0.040, $t(387)$ = -6.488, p < .001, showing that participants were more accurate in answering trials when the second word was in their native language.

An interaction was found between the language of the second word and language congruency, β = 0.032, $t(387)$ = 3.385, p < .001, with a greater difference in accuracy between trials with a language switch when the second word was in English as compared to when it is in Spanish (Figure 17). As with the bilinguals, this is likely due to the differing language skills in both languages: participants are stronger in English than Spanish, and are more likely to answer correctly when the word pair consists of two English words as compared to two Spanish words. Similarly, they are more likely to answer correctly when the trial consists of an English and a Spanish word, as opposed to two Spanish words.

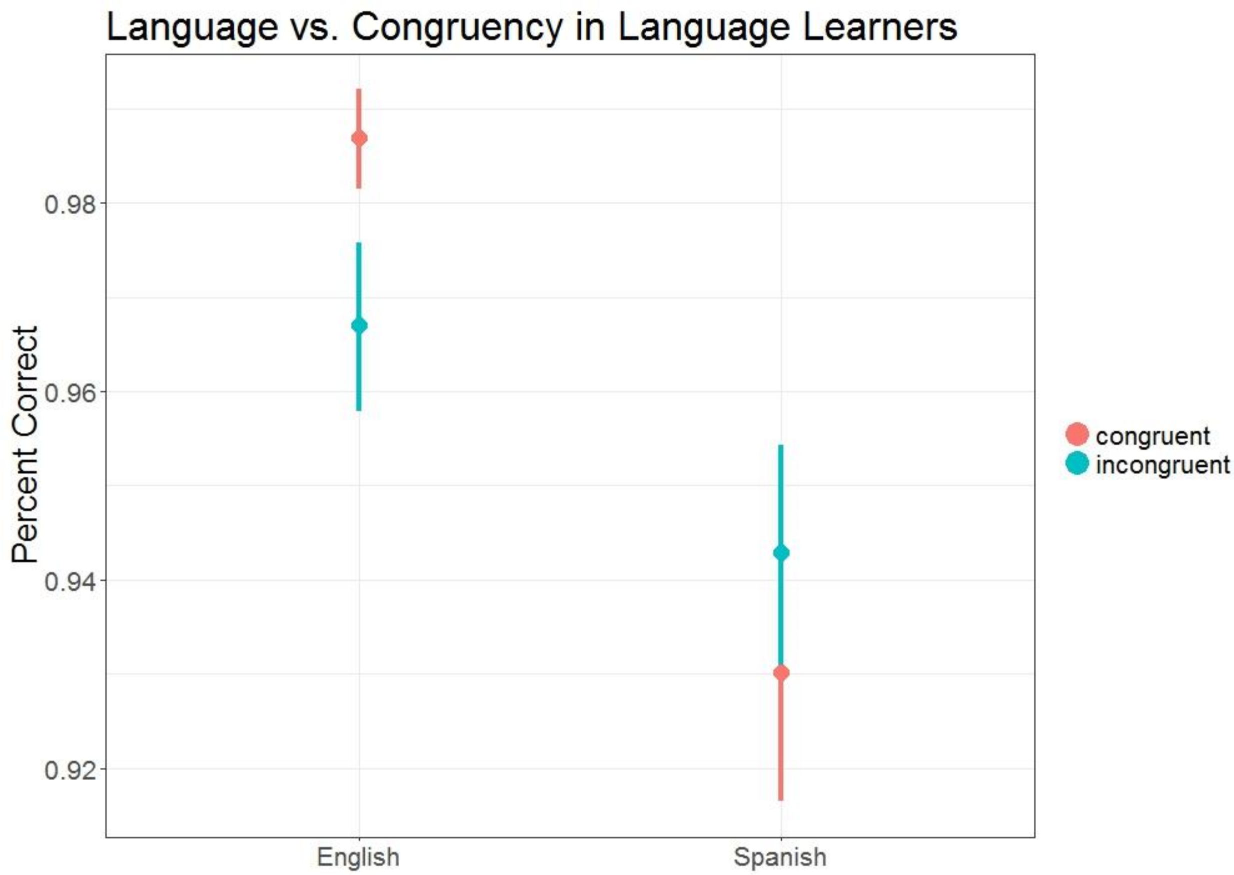


Figure 17: Congruency x language interaction for accuracy analysis in Experiment 2. Language refers to the language of the second word; error bars represent the standard error of the mean.

Spanish proficiency was found to be significant, with accuracy improving as did proficiency, β = 0.008, $t(387) = 2.918$, $p = 0.005$.

Table 10 shows accuracy rates by grammaticality, language condition, and congruency. For each row, the topmost number is the overall percentage correct per condition (not separated by grammaticality), followed by the percentage correct as divided by grammaticality (det-N is grammatical, adv-N is ungrammatical).

Table 10: Average accuracy rates in Experiment 2.

	Percent Correct	
English-English	98.74%	
	det-N	99.67%
	adv-N	97.76%
Spanish-English	96.76%	
	det-N	98.78%
	adv-N	94.64%
English-Spanish	94.32%	
	det-N	94.83%
	adv-N	93.79%
Spanish-Spanish	93.03%	
	det-N	94.30%
	adv-N	91.73%
Same Language	95.87%	
	det-N	96.99%
	adv-N	94.71%
Mixed Language	95.56%	
	det-N	96.82%
	adv-N	94.22%
Word 2 (English)	97.75%	
	det-N	99.22%
	adv-N	96.19%
Word 2 (Spanish)	93.67%	
	det-N	94.56%
	adv-N	92.75%
ALL	95.71%	
	det-N	96.91%
	adv-N	94.46%

Reading time analysis. We fit a linear mixed-effects model to the first pass reading time of the second word of correctly answered 'yes' trials using the lme4 package in R (Bates et al., 2015; R Core Team, 2015). Only trials where the second word had a fixation were included; across the critical trials, nine trials were discarded due to the participant not fixating on the second word (0.13%). Trials without a fixation on the first word were included, which accounted for 227 trials (3.38%). Grammatical trials were 3 times more likely to have the first word skipped than ungrammatical trials. The language of the words nor whether or not there was a language switch did not affect whether or not the first word was skipped.

Analysis not including Spanish proficiency as a moderator for grammaticality. The linear mixed-effects model included the following fixed variables: length and log frequency of the second word, length and log frequency of the first word, trial number, English and Spanish proficiency scores as obtained on the lexTALE and the lexTALE-ESP tests respectively, and our main variables of interest, language of the second word, language congruency, and grammaticality, as well as all interactions of those three variables. All variables were centered before being added to the model.

Random effects were taken into account for both item and subject. As before, we followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for word length of the second word, frequency of the second word, language of the second word, frequency of the first word, language congruency, and trial number; for items, a random intercept, and slopes for language congruency and trial number (Figure 18).

```
Model <- lmer( FPRT ~ TrialNumber  +
                  EnglishProficiency +
                  SpanishProficiency +
                  WordLength +
                  LogWordFrequency +
                  LogSpilloverFrequency +
                  SpilloverLength +
                  LanguageCongruence*Grammaticality*WordLanguage +
                  (1 |subject) +
                    (0 + WordLength |subject) +
                    (0 + LogWordFrequency |subject) +
                    (0 + LogSpilloverFrequency |subject) +
                    (0 + LanguageCongruence |subject) +
                    (0 + WordLanguage |subject) +
                    (0 + TrialNumber |subject) +
                  (1 |item) +
                    (0 + LanguageCongruence |item) +
                    (0 + TrialNumber |item),
                  data=learner, REML=F)
```

Figure 18: Linear mixed-effects model for reading time analysis in Experiment 2.

There was a trend for nouns to be read faster in a grammatical word pair as compared to an ungrammatical word pair, *β* = 8.910, *t*(6692) = 1.758, *p* = 0.08. Though the grammaticality effect did not reach significance, the trend is in the expected direction. Additionally, we found a significant language congruency effect, *β* = 19.495, *t*(6692) = 5.682, *p* < .001, with words read slower following a language switch than when following a word of the same language. We also found a language effect, *β* = 60.935, *t*(6692) = 9.240, *p* < .001. As expected, Spanish words were read more slowly than English words, since our participants were native speakers of English, living in an English-speaking environment. No interactions between our main variables of interest were seen. As no interaction between language and congruency was found, there is no evidence for asymmetrical switch costs; the same switch cost was found regardless of whether the participant switched from their stronger language into their weaker language or vice versa. Figure 19 shows the 95% confidence interval for the main variables of interest, in milliseconds; the dotted line indicates the baseline (English, congruent, grammatical).

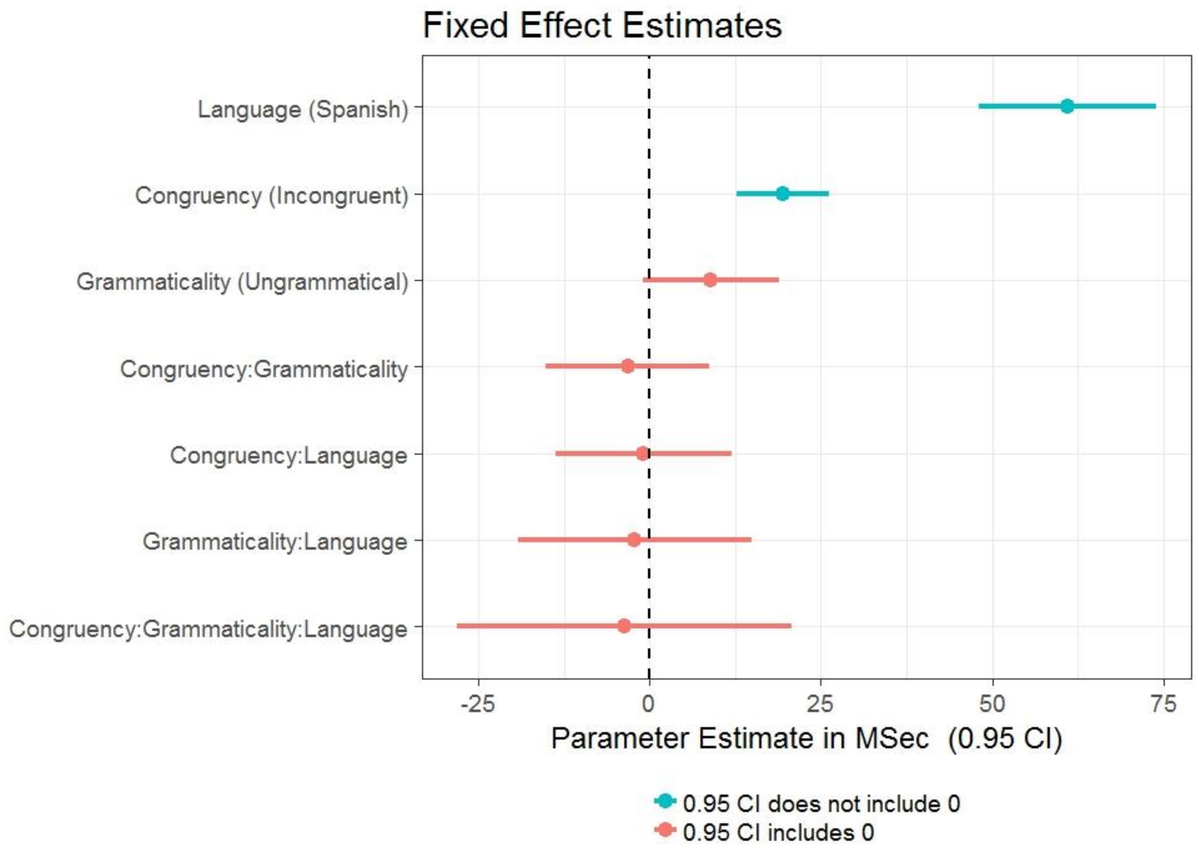


Figure 19: Fixed effect estimates for reading time in Experiment 2.

Of the control variables, we found main effects for the frequency and length of the second word, with more frequent and shorter words being read faster. (Frequency of Word 2: β = -20.953, t(6692) = -6.145, p < .001; Length of Word 2: β = 50.927, t(6692) = 11.965, p < .001) No spillover length or frequency effects were found. We found a practice effect, with participants becoming faster as the experiment went on, β = -20.641, t(6692) = -3.242, p = 0.002. We found no significant proficiency effects for either Spanish or English.

Table 11 shows the results for the fixed variables of the model shown in Figure 18. *P*-values were generated using the lmerTest package (Kuznetsova et al., 2016).

Table 11: Results of the linear mixed-effects model for reading time in Experiment 2. Significant results are bolded.

Variable	β	*t*-value	*p*-value
Intercept	**383.109**	**33.229**	**<.001**
TrialNumber	**-20.641**	**-3.242**	**0.002**
WordFrequency	**-20.953**	**-6.145**	**<.001**
WordLength	**50.927**	**11.965**	**<.001**
SpilloverFrequency	4.211	1.324	0.19
SpilloverLength	3.063	1.060	0.29
EnglishProficiency	-5.886	-0.506	0.62
SpanishProficiency	-9.219	-0.796	0.43
Language [Spanish]	**60.935**	**9.240**	**<.001**
Congruency [Incongruent]	**19.495**	**5.682**	**<.001**
Grammaticality [Ungrammatical]	8.910	1.758	0.08
Congruency*Grammaticality	-3.219	-0.528	0.60
Congruency*Language	-0.897	-0.137	0.89
Grammaticality*Language	-2.147	-0.248	0.80
Congruency*Grammaticality*Language	-3.724	-0.300	0.76

In Figure 20, we show the values of the predicted reading times as calculated using the coefficients shown in Table 11 for our three main variables of interest, language, language congruence, and grammaticality. The trend towards an effect of grammaticality can be seen for all language combinations, with words in grammatical contexts read faster than words in ungrammatical contexts. Words following a language change are read slower than words not following a language change (congruency). This can be seen by comparing the English-English and Spanish-English language pairs, as well as the Spanish-Spanish and English-Spanish language pairs—the trend was seen for both English and Spanish words. Though the predicted times appear to show variability in the effect of grammaticality, interactions were not significant.

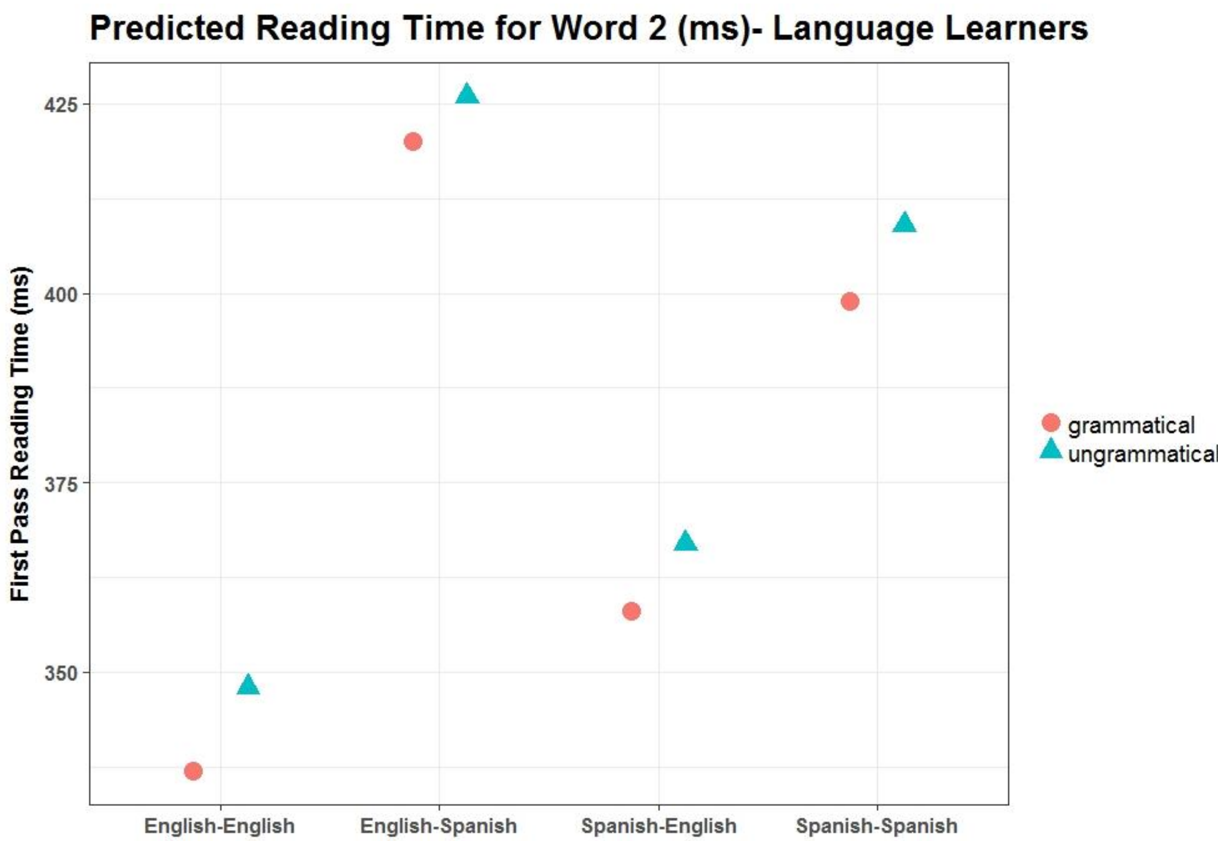


Figure 20: Point estimates for reading times given by linear mixed-effects model in Experiment 2.

Analysis with Spanish proficiency as a moderator for grammaticality. Since we were unsure whether differences in L1 and L2 proficiency would affect whether or not grammaticality effects are found, an additional analysis was performed that included interactions of L2 proficiency with grammaticality, language, and language switch costs. (L1 proficiency was not included as participants were native speakers and had very low variance between participants.) Significant interactions with Spanish proficiency would indicate that differences in L1 and L2 proficiency may modulate effects.

The linear mixed-effects model included the following fixed variables: length and log frequency of the second word, length and log frequency of the first word, trial number, English proficiency scores as obtained on the lexTALE test, and our main variables of interest, Spanish proficiency (scores obtained on the lexTALE-ESP test), language of the second word, language congruency, and grammaticality, as well as all possible interactions of those four variables. All variables were centered before being added to the model.

Random effects were taken into account for both item and subject. As before, we followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for word length of the second word, frequency of the second word, frequency of the first word, language of the second word, language congruency, and trial number; for items, a random intercept, and slopes for language congruency and trial number (Figure 21).

```
Model <- lmer( FPRT ~ TrialNumber  +
                   EnglishProficiency +
                   WordLength +
                   LogWordFrequency +
                   LogSpilloverFrequency +
                   SpilloverLength +
                   SpanishProficiency*LanguageCongruence*Grammaticality*WordLanguage +
                   (1 |subject) +
                     (0 + WordLength |subject) +
                     (0 + LogWordFrequency |subject) +
                     (0 + LogSpilloverFrequency |subject) +
                     (0 + LanguageCongruence |subject) +
                     (0 + WordLanguage |subject) +
                     (0 + TrialNumber |subject) +
                   (1 |item) +
                     (0 + LanguageCongruence |item) +
                     (0 + TrialNumber |item),
                   data=learner, REML=F)
```

Figure 21: Linear mixed-effects model for reading time analysis including interactions with Spanish proficiency in Experiment 2.

There was a trend for nouns to be read faster in a grammatical word pair as compared to an ungrammatical word pair, $\beta = 8.777$, $t(6685) = 1.733$, $p = 0.08$. Though the grammaticality effect did not reach significance, the trend is in the expected direction. Spanish proficiency was not significant, but there we found a three-way interaction between congruency, grammaticality, and Spanish proficiency, $\beta = -16.227$, $t(6685) = -2.755$, $p = 0.006$, as can be seen in Figure 22. The interaction is likely spurious, as it suggests language switch cost changes depending on grammaticality. The ungrammatical word pairs show the most likely pattern, based on previous research; the most proficient learners show the least switch cost. However, this pattern is reversed when looking at the grammatical pairs, with the least proficient learners showing less switch cost than more proficient learners, an unlikely finding, especially since there are no interactions between proficiency and congruency, nor proficiency and grammaticality.

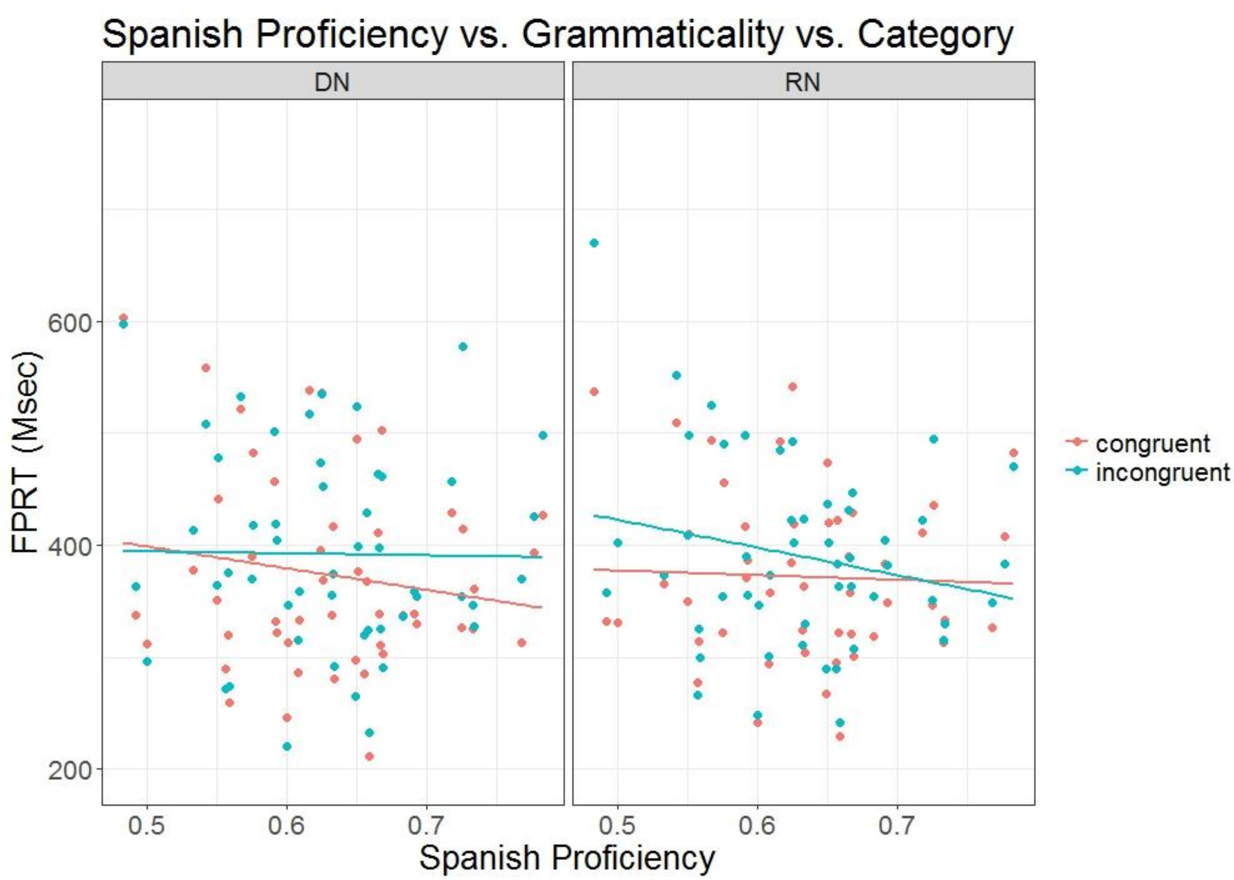


Figure 22: Spanish proficiency x congruency x grammaticality interaction for reading time analysis including interactions with Spanish proficiency in Experiment 2.

We also found a significant language congruency effect, $\beta = 19.491$, $t(6685) = 5.691$, $p < .001$, with words read slower following a language switch than when following a word of the same language. We also found a language effect, $\beta = 60.835$, $t(6685) = 9.368$, $p < .001$. As expected, Spanish words were read more slowly than English words, since our participants were native speakers of English, living in an English-speaking environment. Interestingly, language did not interact with Spanish proficiency, suggesting that less proficient speakers did not show worse performance on Spanish words than more proficient speakers.

No other interactions between our main variables of interest were found, including no evidence for asymmetrical language switch cost. Figure 23 shows the 95% confidence interval for the main variables of interest, in milliseconds; the dotted line indicates the baseline (English, congruent, grammatical).

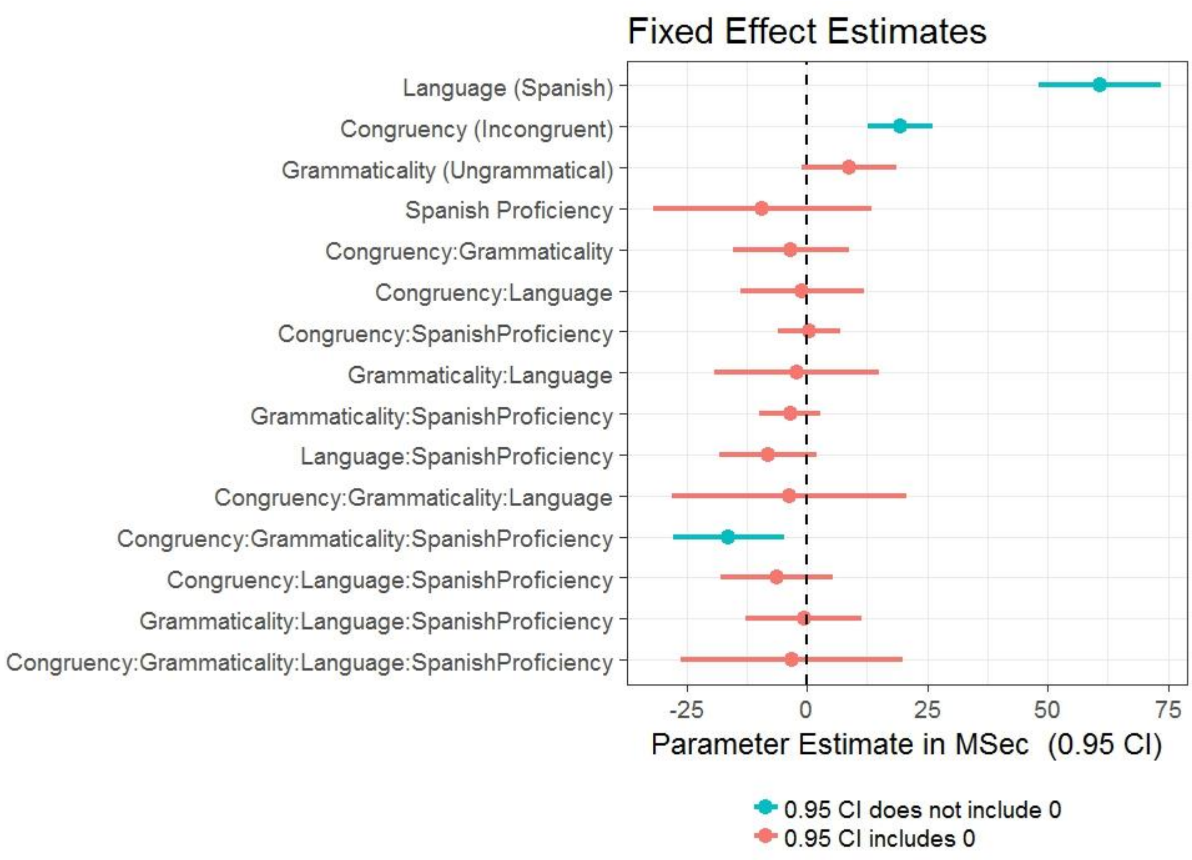


Figure 23: Fixed effect estimates for reading time analysis including interactions with Spanish proficiency in Experiment 2.

Of the control variables, we found main effects for the frequency and length of the second word, with more frequent and shorter words being read faster. (Frequency of Word 2: $\beta = -20.919$, $t(6685) = -6.155$, $p < .001$; Length of Word 2: $\beta = 50.871$, $t(6685) = 11.950$, $p < .001$) No spillover effects were found. We found a practice effect, with participants becoming faster as the experiment went on, $\beta = -20.756$, $t(6685) = -3.257$, $p = 0.002$. We found no significant English proficiency effect.

Table 12 shows the results for the fixed variables of the model shown in Figure 21. *P*-values were generated using the lmerTest package (Kuznetsova et al., 2016).

Table 12: Results of the linear mixed-effects model for reading time including interactions with Spanish proficiency in Experiment 2. Significant results are bolded.

Variable	*β*	*t*-value	*p*-value
Intercept	**383.064**	**33.253**	**<.001**
TrialNumber	**-20.756**	**-3.257**	**0.002**
WordFrequency	**-20.919**	**-6.155**	**<.001**
WordLength	**50.871**	**11.950**	**<.001**
SpilloverFrequency	4.188	1.317	0.19
SpilloverLength	3.142	1.088	0.28
EnglishProficiency	-5.867	-0.505	0.62
Language [Spanish]	**60.835**	**9.368**	**<.001**
Congruency [Incongruent]	**19.491**	**5.691**	**<.001**
Grammaticality [Ungrammatical]	8.777	1.733	0.08
SpanishProficiency	-9.317	-0.805	0.42
Congruency*Grammaticality	-3.278	-0.539	0.59
Congruency*Language	-0.966	-0.148	0.88
Congruency*SpanishProficiency	0.612	0.185	0.85

Grammaticality*Language	-2.088	-0.241	0.81
Grammaticality*SpanishProficiency	-3.508	-1.091	0.28
Language*SpanishProficiency	-8.030	-1.553	0.13
Congruency*Grammaticality*Language	-3.712	-0.299	0.76
Congruency*Grammaticality*SpanishProficiency	**-16.227**	**-2.755**	**0.006**
Congruency*Language*SpanishProficiency	-6.212	-1.051	0.29
Grammaticality*Language*SpanishProficiency	-0.672	-0.110	0.91
Congruency*Grammaticality*Language*SpanishProficiency	-3.025	-0.257	0.80

Overall, both the model including Spanish proficiency interacting with the variables of interest (Figure 21) and the model not including the Spanish proficiency interaction (Figure 18) show similar patterns among the factors of interest. The biggest difference between the models is that a three-way interaction between Spanish proficiency, grammaticality, and language congruency was found. The interaction is most likely spurious, as it indicates that language congruency is moderated by grammaticality.

Probability of an eye movement regression analysis. 5.70% of accurate trials included a regression from the second word back to the first. To see if these regressions affected the results of the reading time analysis, we looked at the probability that a trial included an eye movement regression back to the first string of characters. Table 13 shows the percent of trials with an eye movement regression as separated by language, language congruency, and grammaticality.

Table 13: Probability of an eye movement regression in Experiment 2.

	Exp 2: Language Learners	
English-English	3.88%	
	det-N	3.78%

	adv-N	3.98%
Spanish-English	7.17%	
	det-N	7.98%
	adv-N	6.28%
English-Spanish	5.96%	
	det-N	6.16%
	adv-N	5.74%
Spanish-Spanish	5.85%	
	det-N	6.64%
	adv-N	5.01%
Same Language	4.84%	
	det-N	5.17%
	adv-N	4.48%
Mixed Language	6.57%	
	det-N	7.09%
	adv-N	6.01%
Word 2 (English)	5.51%	
	det-N	5.87%
	adv-N	5.12%
Word 2 (Spanish)	5.90%	
	det-N	6.40%
	adv-N	5.38%
ALL	5.70%	
	det-N	6.13%
	adv-N	5.25%

One pattern that is evident for all language conditions is that eye movement regressions are more common among grammatical trials as compared to ungrammatical trials. However, as mentioned when looking at Experiment 1, the different word properties of determiners and adverbs may be driving the pattern; it may not be the result of grammaticality. Determiners are both shorter and more frequent than adverbs on average, leading to a correlation between grammaticality and Word 1 properties (between grammaticality and Word 1 frequency: $r(1142) = -0.42$, $p < .001$; between grammaticality and Word 1 length: $r(1142) = 0.55$, $p < .001$).

To analyze the data, we performed a mixed-effects logistic regression using whether or not a participant made an eye movement regression as our dependent variable. Trials with a regression were considered hits, and trials without a regression were considered misses. As in the previous analyses, the fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, English and Spanish proficiency scores as obtained on the lexTALE and the lexTALE-ESP tests respectively, and our main variables of interest: language of the second word, language congruency, and grammaticality, as well as all interactions of those three variables. All variables were centered before being added to the model.

Random effects were taken into account for both item and subject. We again followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for trial number, length of the second word, length of the first word, and frequency of the first word; for items, a random intercept, and slopes for trial number, English proficiency, language congruency, and grammaticality (Figure 24).

```
Model <- glmer(EyeRegression ~ TrialNumber +
                        EnglishProficiency +
                        SpanishProficiency +
                        WordFrequency +
                        WordLength +
                        SpilloverFrequency +
                        SpilloverLength +
                        WordLanguage*LanguageCongruence*Grammaticality +
                        (1 |subject) +
                          (0 + WordLength |subject) +
                          (0 + SpilloverFrequency |subject) +
                          (0 + SpilloverLength |subject) +
                          (0 + TrialNumber |subject) +
                        (1 |item) +
                          (0 + EnglishProficiency |item) +
                          (0 + LanguageCongruence |item) +
                          (0 + Grammacality |item) +
                          (0 + TrialNumber |item),
                      data=learner, family= binomial(link = 'logit'),
                      control=glmerControl(optimizer = 'bobyqa',
                      optCtrl=list(maxfun=100000)))
```

Figure 24: Logistic mixed-effects model for probability of an eye movement regression in Experiment 2.

Mixed language trials were more likely to have an eye movement regression than same language trials, $\beta = 0.365$, $t(6692) = 2.983$, $p = 0.003$, showing a language congruency effect. A main effect of trial number was found, with participants being more likely to make a regression early in the experiment as compared to later in the experiment, $\beta = -0.810$, $t(6692) = -5.090$, $p < .001$. A main effect of length of Word 1 was found, with shorter Word 1 words being more likely to be regressed to, $\beta = -0.283$, $t(6692) = -2.569$, $p = 0.01$. A main effect of length of Word 2 was also found, with participants being more likely to regress the longer Word 2 is, $\beta = 0.154$, $t(6692) = 2.091$, $p = 0.04$. Additionally, a main effect of frequency of Word 2 was also found, with participants being less likely to regress back to Word 1 as the frequency of Word 2 increases, $\beta = -0.134$, $t(6692) = -2.101$, $p = 0.04$.

No main effect of language was found. No main effect of grammaticality was found, despite the fact that the numeric pattern is seen across all language combinations (Table 13). This is likely due to the fact that length of Word 1 was significant. An interaction was found between congruency and language, $\beta = -0.866$, $t(6692) = -3.429$, $p = 0.001$ (Figure 25). Interestingly, different patterns were found for English and Spanish. When the second word is in English there are less regressions in English-English trials as compared to Spanish-English trials (~7% as compared to ~4%). When the second word is in Spanish, however, there are equal numbers of regressions regardless of the language of the first word (~6%).

This interaction is likely the result of several factors. First, as we previously mentioned, a general effect of congruency is present, indicating a general language switching effect—participants are more likely to regress on mixed language trials than same language trials. Second, it likely indicates higher difficulty in processing Spanish words as compared to English words; previous research has shown that the probability of eye movement regressions increases as difficulty increases (Rayner et al., 2006). These two factors interact to cause the effect seen above. When the second word is in English, due to the language of the words, Spanish-English trials are expected to be more difficult than English-English trials; they are also predicted to be so because Spanish-English trials contain a language switch but English-English trials do not. When the second word is in Spanish, however, due to the language of the two words, English-Spanish trials are expected to be easier than Spanish-Spanish trials (only one Spanish word as compared to two); at the same time, English-Spanish trials are expected to be harder than Spanish-Spanish trials because processing difficulty increases following a language switch. The two factors are affecting the data in opposing ways, causing there to be no difference in processing cost between the two types of trial. The data also follow the predicted pattern of differences in switch cost, with larger switch costs for L2 -> L1 (Spanish-English) than L1 -> L2 (English-Spanish). Though tempting to assume that the interaction shows a difference in switch cost, it is unlikely, as the difference in eye movement regression probability between L1 -> L2 and L2 -> L1 falls well within the error bars for the points.

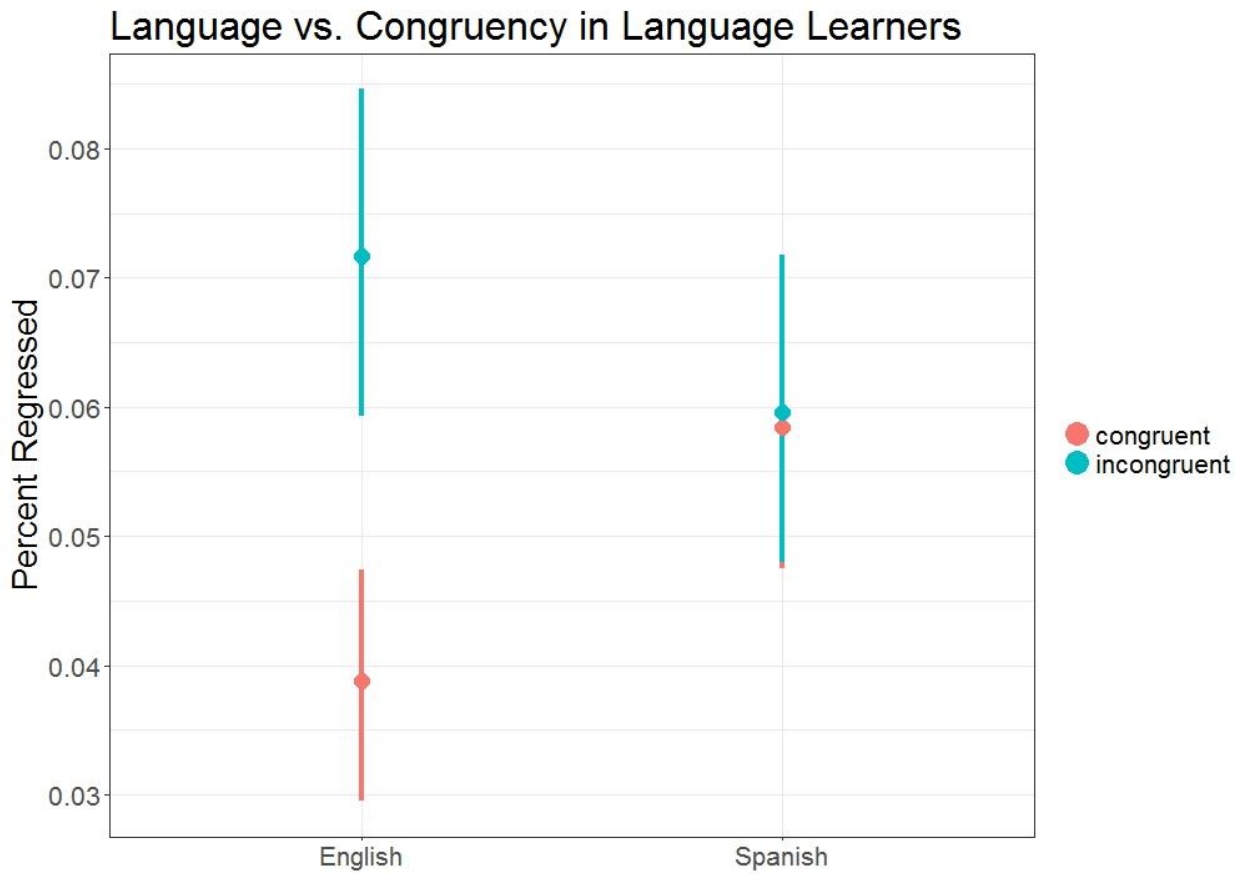


Figure 25: Congruency x language interaction for probability of an eye movement regression in Experiment 2. Language refers to the language of the second word; error bars represent the standard error of the mean.

Reading time analysis for trials without an eye movement regression. No main effect of grammaticality was found in the probability of eye movement regression logistic regression. However, it is possible that the reason that it does not come out in the model is due to collinearity between grammaticality and Word 1 properties (frequency and length), so we cannot definitively state that grammaticality is not seen in eye movement regressions. Though grammaticality and Word 1 frequency and length remain variables in the linear regression, by removing trials with eye movement regressions we remove noise due to such regressions. We can say that we are removing outlier trials of a sort, ones that do not have the most typical reading behavior. We ran a linear mixed-effects regression using first pass reading time as the dependent variable only on trials without a regression.

As in the previous analyses, the fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, English and Spanish proficiency scores as obtained on the lexTALE and the lexTALE-ESP tests respectively, and our main variables of interest: language of the second word, language congruency, and grammaticality, as well as all interactions of those three variables. All variables were centered before being added to the model.

Random effects were taken into account for both item and subject. We again followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for trial number, word length of the second word, frequency of the second word, and language of the second word; for items, a random intercept, and slopes for trial number, language congruency, and grammaticality (Figure 26).

```
Model <- lmer(FPRT ~ TrialNumber +
                     EnglishProficiency +
                     SpanishProficiency +
                     WordFrequency +
                     WordLength +
                     SpilloverFrequency +
                     SpilloverLength +
                     WordLanguage*LanguageCongruency*Grammaticality +
                     (1 |subject) +
                       (0 + WordFrequency |subject) +
                       (0 + WordLength |subject) +
                       (0 + WordLanguage |subject) +
                       (0 + TrialNumber |subject) +
                     (1 |item) +
                       (0 + LanguageCongruency |item) +
                       (0 + Grammaticality |item) +
                       (0 + TrialNumber |item),
                   data=learner[no regressions], REML=F)
```

Figure 26: Linear mixed-effects model for reading time analysis excluding trials with an eye movement regression in Experiment 2.

As stated, the main reason to look at this model was to see if the main effect of grammaticality would become statistically significant if the trials without regressions were included. Unlike in Experiment 1, no effect of grammaticality was found, suggesting that the eye-movement regressions may be in part driving the observed trend. This is perhaps not surprising, given that grammatical trials have higher rates of eye movement regressions.

Mixed language trials were read more slowly than same language trials, $\beta = 20.371$, $t(6310) = 6.801$, $p < .001$, showing a language switch cost. Trials with the second word in Spanish were read more slowly than those with the second word in English, $\beta = 61.915$, $t(6310) = 9.035$, $p < .001$. No interactions among the variables of interest were found.

In regards to the variables controlling for extra factors, similar effects were found regarding word length and word frequency. Reading time increased when the length of the second word increased, $\beta = 53.144$, $t(6310) = 11.943$, $p < .001$, and when the frequency of the second word decreased, $\beta = -21.219$, $t(6310) = -5.955$, $p < .001$. A main effect of trial number was also found, showing a practice effect, as participants became faster as the experiment went on, $\beta = -28.013$, $t(6310) = -4.261$, $p < .001$.

No proficiency effects were found for either Spanish or English. A main effect of frequency of Word 1 was found, with more frequent Word 1 words being associated with longer reading times of Word 2, $\beta = 5.271$, $t(6310) = 2.030$, $p = 0.04$. This is thought to be due to imperfect stimuli used in the study, as words following adverbs were more frequent than words following determiners. Table 14 shows the results for the fixed variables of the model shown in Figure 26.

Table 14: Results of the linear mixed-effects model excluding trials with an eye movement regression in Experiment 2. Significant results are bolded.

Variable	β	*t*-value	*p*-value
Intercept	**387.240**	**32.809**	**<.001**
TrialNumber	**-28.013**	**-4.261**	**<.001**
WordFrequency	**-21.219**	**-5.955**	**<.001**
WordLength	**53.144**	**11.943**	**<.001**

SpilloverFrequency	**5.271**	**2.030**	**0.04**
SpilloverLength	1.710	0.598	0.55
EnglishProficiency	-6.521	-0.547	0.59
SpanishProficiency	-10.114	-0.856	0.40
Language [Spanish]	**61.915**	**9.039**	**<.001**
Congruency [Incongruent]	**20.371**	**6.801**	**<.001**
Grammaticality [Ungrammatical]	6.729	1.302	0.19
Congruency*Grammaticality	-5.645	-0.953	0.34
Congruency*Language	-2.046	-0.322	0.75
Grammaticality*Language	-3.930	-0.443	0.66
Congruency*Grammaticality*Language	0.067	0.006	0.99

Reading time analysis comparing grammaticality in same and mixed language trials. In the previous reading time analyses, we found a grammatical predictability effect, and no significant interaction between grammaticality and congruency. As in Experiment 1, however, if we wish to state that grammatical predictability is language-independent, we must compare the effect size of grammaticality for both same language and mixed language trials. To do so, we separated grammaticality into two variables instead of looking for an interaction between grammaticality and congruency; grammaticality in congruent trials, and grammaticality in congruent trials.

We fit a linear mixed-effects model to the first pass reading time of the second word of correctly answered critical 'yes' trials using the lme4 package in R (Bates et al., 2015; R Core Team, 2015). As in the reading time analysis, the fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, English and Spanish proficiency scores as obtained on the lexTALE and the lexTALE-ESP tests respectively, and our main variables of interest: language of the second word, language congruency, grammaticality in congruent trials, and grammaticality in incongruent trials. Interactions between language and congruency, language and grammaticality in congruent trials, and grammaticality in incongruent trials were also included. All variables were centered before being added to the model.

Random effects were taken into account for both subject and item. Subject referred to the participant; item referred to the second word. Each participant was exposed to an item at most once in the experiment (participants were not exposed to all items, due to each word pair appearing in one of four possible language conditions, of which a participant was exposed to one). As in the previous reading time analysis, we followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for word length of the second word, word frequency of the second word, word frequency of the first word, language of the second word, language congruency, and trial number; for items, a random intercept, and slopes for language congruency, grammaticality in congruent trials, grammaticality in incongruent trials, and trial number (Figure 27).

```
Model <- lmer(FPRT ~ TrialNumber  +
                  EnglishProficiency +
                  SpanishProficiency +
                  WordLength +
                  LogWordFrequency +
                  LogSpilloverFrequency +
                  SpilloverLength +
                  CongruentGrammaticality*WordLanguage +
                  IncongruentGrammaticality*WordLanguage +
                  LanguageCongruence*WordLanguage +
                  (1 |subject) +
                    (0 + WordLength |subject) +
                    (0 + LogWordFrequency |subject) +
                    (0 + LogSpilloverFrequency |subject) +
                    (0 + LanguageCongruence |subject) +
                    (0 + WordLanguage |subject) +
                    (0 + TrialNumber |subject) +
                  (1 |item) +
                    (0 + LanguageCongruence |item) +
                    (0 + CongruentGrammaticality |item) +
                    (0 + IncongruentGrammaticality |item) +
                    (0 + TrialNumber |item),
                  data=learner, REML=F)
```

Figure 27: Linear mixed-effects model for reading time analysis comparing grammaticality in same and mixed language trials in Experiment 2.

Of our grammaticality variables, we found a trend for grammatical predictability in congruent trials, $\beta = 10.501$, $t(6690) = 1.728$, $p = 0.08$, with nouns being read faster in grammatical same language word pairs as compared to ungrammatical same language word pairs, but it did not reach significance. Grammaticality in incongruent trials had a slightly lower magnitude, $\beta = 7.308$, $t(6690) = 1.230$, $p = 0.22$, but the pattern was similar for both types of word pairs. Grammaticality was not found to interact with language in either grammatical or ungrammatical trials.

As in previous reading time analyses, we found a language effect, $\beta = 60.858$, $t(6690) = 9.229$, $p < .001$, with Spanish words being read more slowly than English words. We also found an effect of language congruency, $\beta = 19.467$, $t(6690) = 5.706$, $p < .001$, showing that nouns following a language switch were read slower than nouns following a word of the same language. No interaction was found between language and congruency. Figure 28 shows the 95% confidence interval for the main variables of interest, in milliseconds; the dotted line indicates the baseline (English, congruent, grammatical).

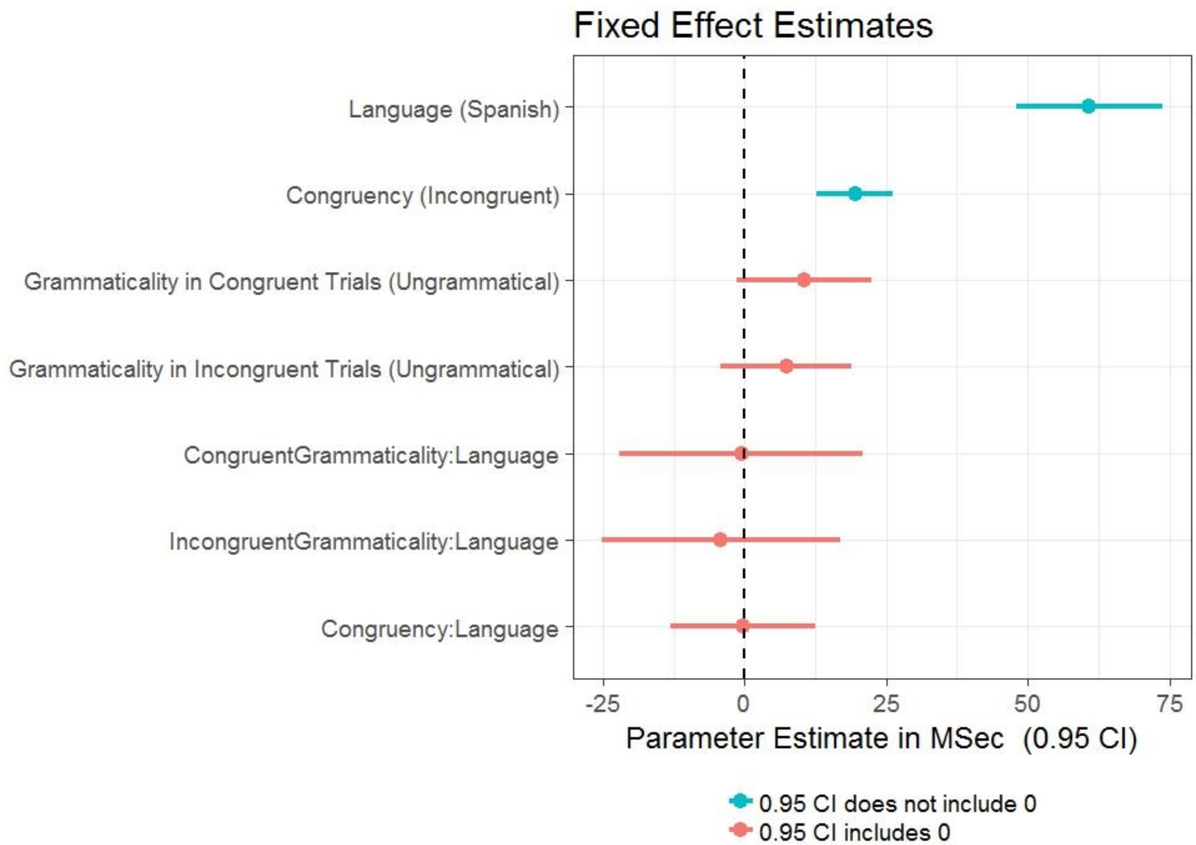


Figure 28: Fixed effect estimates for reading time analysis comparing grammaticality in same and mixed language trials in Experiment 2.

We found main effects for the frequency and length of the second word, with more frequent and shorter words being read faster. (Frequency of Word 2: $\beta = -20.976$, $t(6690) = -6.153$, $p < .001$; Length of Word 2: $\beta = 50.934$, $t(6690) = 11.944$, $p < .001$) There was a practice effect, $\beta = -20.552$, $t(6690) = -3.221$, $p = 0.002$, with participants becoming faster as the experiment progressed. We did not find any spillover effects, and neither Spanish nor English proficiency were significant.

Table 15 shows the results for the fixed variables of the model shown in Figure 27. *P*-values were generated using the lmerTest package (Kuznetsova et al., 2016).

Table 15: Results of the linear mixed-effects model for reading time analysis comparing grammaticality in same and mixed language trials in Experiment 2. Significant results are bolded.

Variable	*β*	*t*-value	*p*-value
Intercept	**383.167**	**33.235**	**<.001**
TrialNumber	**-20.552**	**-3.221**	**0.002**
WordFrequency	**-20.976**	**-6.153**	**<.001**
WordLength	**50.934**	**11.944**	**<.001**
SpilloverFrequency	4.267	1.338	0.18
SpilloverLength	3.156	1.087	0.28
EnglishProficiency	-5.885	-0.506	0.62
SpanishProficiency	-9.237	-0.798	0.43
Language [Spanish]	**60.858**	**9.229**	**<.001**
Congruency [Incongruent]	**19.467**	**5.706**	**<.001**
CongruentGrammaticality [Ungrammatical]	10.501	1.728	0.08
IncongruentGrammaticality [Ungrammatical]	7.308	1.230	0.22
CongruentGrammaticality*Language	-0.588	-0.054	0.96
IncongruentGrammaticality*Language	-4.215	-0.392	0.70
Congruency*Language	-0.325	-0.050	0.96

Discussion

Lexical Decision Task

Experiment 2 looked at how proficiency interacted with grammatical predictability. We looked at advanced language learners of Spanish (L1 English), and tested them with the same stimuli and methodology used in Experiment 1. We were interested first in seeing whether language learners showed evidence of grammaticality predictability, second in seeing whether grammatical predictability interacted with language (learners showing grammaticality effects only in their L1), and finally in seeing how they performed in word pairs involving a language shift. Previous research showed that grammaticality effects are found in monolinguals (Staub & Clifton, 2006; Vigliocco et al., 2008), and in Experiment 1, we found evidence of grammatical predictability regardless of language in Spanish-English bilinguals.

We found a strong trend for grammatical predictability in reading times, with participants reading nouns in grammatical language pairs faster than nouns in ungrammatical language pairs. There was no interaction with either language or congruency, suggesting that participants behaved similarly regardless of the language of the noun being read and that grammatical predictability is equally present (or absent) in all language conditions. Looking at grammaticality separately for congruent and incongruent trials shows that the magnitude of the grammatical predictability found was similar for both same and mixed language pairs. We are unable to conclude that this trend provides strong evidence for grammatical predictability, however, as it did not appear in all analyses. More robust grammaticality effects were found by looking at accuracy of response.

In regards to the reading time analysis, we found no interactions between language of the second word, language congruency, and the grammaticality of the word pair. We had expected to find an interaction between language and congruency, due to previous studies finding asymmetrical switch costs (Meuter & Allport, 1999). In other words, we expected that participants would show a larger switch cost when word pairs were in the Spanish-English condition as compared to word pairs in the English-Spanish condition. No such interaction was found in reading times, but we did find an interaction between language and language congruency in both accuracy of response and probability of an eye movement regression, which may indicate a larger switch cost from Spanish to English than from English to Spanish.

We are reluctant to conclude that our data provide evidence of asymmetrical costs in language switching, due to several reasons. First, the difference in switch cost falls well within the error bars of the conditions when looking only at the mixed language conditions. Second, in previous research, asymmetrical switch costs are most often found the greater the difference in proficiency (Costa et al., 2006); our participants were advanced students, taking courses in Spanish, so while they were weaker in Spanish than English, they are still quite proficient. Finally, to our knowledge, asymmetrical language costs have only been found in naming tasks, and not all attempts at replicating the effect have succeeded (Bobb & Wodniecka, 2013). Though classically believed to be caused by greater difficulty in inhibiting a native language as opposed to an L2, other researchers have argued for it to be due to other factors such as persistent activation (Philipp et al., 2007). Due to all these factors, it is more likely that the differences in switch cost found in the accuracy and eye movement regression analyses are due to differences in processing words in different languages.

Participants read words in Spanish slower than words in English, indicating that they find processing English words easier than Spanish word. This is likely due to both greater practice at reading English and greater linguistic proficiency in English as compared to Spanish. Similarly, participants were also more accurate when the noun was in English as compared to in Spanish. As all participants scored at least 30% higher on the English proficiency test (lexTALE) than the Spanish proficiency test (lexTALE-ESP), this finding was expected.

Interestingly, regardless of the fact that participants were stronger in English than in Spanish, there was a lack of interaction between grammaticality and language in all analyses. This shows that participants are behaving similarly in regards to grammatical predictability, no matter the language of the word being read. This is a bit surprising, as participants are more practiced in English than Spanish; if grammatical predictability was due to previous experience with the specific tokens being analyzed we might find greater grammaticality effects in English than in Spanish.

Spanish proficiency and predictability. We were interested in seeing whether Spanish proficiency affected grammatical predictability. Spanish proficiency was not found to be significant, and the linear regression model including an interaction between proficiency and grammaticality was not found to be significantly different than the model not including it. Overall, taken with the fact that Spanish proficiency was not correlated with sensitivity to Spanish-English code-switching constraints, this suggests that all our participants had acquired a level of Spanish linguistic ability high enough that it leads to minimal differences in grammatical predictability.

Language Assessment Tasks

All participants were tested on a battery of language assessment tasks focused on measuring proficiency in Spanish. We used the lexTALE-ESP, an untimed lexical decision task, as our main measure of proficiency in our statistical models; however, we also measured participants using a grammar-based test and using the Spanish EIT, an oral fluency measure. All three scores were correlated, showing that participants with high proficiency tended to score highly in all three dimensions.

In addition to Spanish proficiency measures, participants were tested on a code-switching sentence judgment task, in which they were asked to rate some sentences that were grammatical in that they followed Spanish-English code-switching constraints, and other sentences that were ungrammatical in that they did not conform to those constraints. Participants rated grammatical sentences higher than ungrammatical sentences, suggesting that they were sensitive to Spanish-English constraints. This finding falls in line with previous research that had found that advanced language learners were more sensitive to constraints than more beginner language learners (Toribio, 2001b). However, unlike in the Toribio study, the sensitivity to code-switching constraints was not correlated with Spanish proficiency (by any measure).

Limitations

Our results were inconclusive, with grammaticality effects being found in accuracy analyses, but only unreliable trends in reading time analyses. Some patterns were stable, however; where we found a grammaticality effect or a trend of grammaticality, there was no interaction with language congruency. This suggests that language learners show language-independent grammatical predictability effects when found.

As we wanted to be most likely to observe grammaticality effects, we looked only at very high-proficiency language learners. This is not representative of all language learners, especially since not all models of language learning posit that beginner language learners necessarily represent languages the same as more advanced learners. Therefore, we cannot draw conclusions as to the behavior of language learners in general; instead, we looked only at a small, select sample of the population. A different sample of less proficient learners would need to be tested to see whether or not they show language-independent grammatical predictability.

Additionally, we are limited by the participants we recruited, and their current circumstances. All participants were currently in the United States, and being exposed to Spanish (and Spanish-English code-switching) primarily in the classroom, in which code-switching is usually seen negatively, used when a language learner cannot think of the word in the L2. It would be interesting to look at a language learner population which uses code-switching more often in more natural settings; some such populations may be recent immigrants in an immersion setting (such as some bilingualism transition programs, as can be found in schools with large numbers of immigrants), or college students studying abroad, where they are interacting with speakers of the L2 in a more casual setting. Though some of our participants had studied abroad, all currently lived in the United States and used English more often than Spanish.

Another issue is the fact that word frequency has strong effects on reading times, and so we needed to take it into account in our statistical models. The concern lies with the fact that Spanish word frequencies are drawn from Spanish language sources, but that the words language learners are exposed to do not necessarily fall in the same frequency patterns. In other words, some words occur more often in normal Spanish sources than in Spanish textbooks, while others occur less often. In particular, abstract concepts (e.g., 'tema' - topic) tend to be under-represented, but concrete nouns tend to be over-represented (e.g., 'calabaza' - pumpkin) (Davies & Face, 2006). This is especially relevant since our stimuli were carefully chosen to be words encountered in the first semester of Spanish language learning (as much as possible; due to the fact that we had to use a certain number of adverbs, which tend not to occur too often in textbooks, some came from frequency lists).

Though a concern, the fact that we used advanced language learners makes it more likely that they had been exposed to more natural vocabulary patterns than just occurring in the Spanish textbook, as they read Spanish books are were exposed to Spanish media. However, we cannot assume that the Spanish frequency values are altogether accurate for the language learners, and this likely adds noise to the data.

Conclusions

In summary, our study suggests that advanced English-Spanish language learners are behaving similarly to Spanish-English bilinguals. However, we were unable to make definitive conclusions based solely on the language learner data, and would have to compare the data directly to draw stronger conclusions. As both groups of participants were tested using the same methodology and stimuli, we are able to analyze the data jointly, which would allow us to determine whether or not advanced English-Spanish language learners are behaving as do Spanish-English bilinguals.

Additionally, though we cannot draw conclusions as to the effect of L2 proficiency on grammatical predictability (as seen by looking at advanced language learners), other factors may also be affecting grammatical predictability. One possibility is code-switching can be thought of as having its own separate grammar, rather than being simply the intersection between two languages. Different pairs of languages have different code-switching constraints (MacSwan, 2000), with some allowing constructions that are not part of either of the two languages (Bentahila & Davies, 1983; Berk-Seligson, 1986). Learning such constraints may be akin to learning the grammar of a new language. If this is the case, we might expect to see grammatical predictability effects to be modulated by sensitivity to code-switching constraints, or by experience with code-switching.

In the next chapter, we analyze the data from Experiments 1 and 2 jointly, looking for whether or not language learners show the same grammatical predictability effects as do bilinguals. Moreover, we also look at whether sensitivity to code-switching constraints or code-switching use affect grammatical predictability.

Chapter 4

Comparisons between Experiments 1 and 2

Introduction

How similar bilinguals and language learners are is a question under debate. Experiments 1 and 2 used the same paradigm to test two different groups: Spanish-English bilinguals, and advanced language learners of Spanish. In Experiment 1, we found evidence suggesting that Spanish-English bilinguals showed language-independent grammatical predictability in reading times, but not when looking at accuracy of response. In Experiment 2, we found evidence that advanced English-Spanish language learners showed language-independent grammatical predictability in accuracy of response. However, results were not significant when looking at reading times, not allowing us to conclude that they showed grammatical predictability in reading times.

In this chapter, we wish to compare the participants from Experiment 1 and 2 directly, so as to draw more definitive conclusions. Additionally, we wish to look at the effect on grammatical predictability of other variables, to determine whether or not factors other than proficiency may be affecting grammatical predictability.

There are several reasons why language learners might perform differently as compared to bilinguals, related to either their quality of language knowledge, or their language use. One reason that language learners might perform differently to bilinguals is that they are only 'native' in one of the languages—Spanish was learned past childhood. One common theory is that a language must be learned in early childhood for a speaker to have full proficiency; that a person passes through a certain stage of development which does not allow for attainment to reach as high a level—the critical period hypothesis (Johnson & Newport, 1989; Lenneberg, 1967). In this dissertation, we shall not go into detail on the theory, as it does not pertain to whether or not bilinguals and language learners have shared representations of their languages. However, one potential implication is that it may imply that a native language and a second language are represented and processed differently in the brain (Ullman, 2001), regardless of the participants' proficiency in the languages.

Another possibility is that participants' different experiences with language use and code-switching may lead to differences in processing code-switching sentences. Even small differences in relative frequency can lead to differences in reading time for a word, especially if the words are infrequent (Smith & Levy, 2013). We are looking at code-switches, which occur with very low frequency even among communities that code-switch often (Poplack, 1980). While we asked participants to rate the amount of time spent code-switching, these are self-reported measures, which are known to be influenced by personality, age of acquisition of L2, as well as living and working situations (Dewaele & Li, 2014). Bilinguals and language learners might be using different internal measures to estimate the amount of time spent code-switching. Even if they ultimately do not code-switch more often, bilingual participants are expected to code-switch in wider circumstances than language learners, because they are assumed to be more likely to use the language in a variety of ways, such as with friends and family, rather than primarily in select classes. This difference in quality of use might also lead to differences in processing.

Comparing Experiments 1 and 2

There are several questions we wished to focus upon by comparing the data gathered in Experiments 1 and 2. We are primarily interested in seeing whether type of participant (bilingual or language learner) is a predictor of behavior in the experiment, both in reading times and in accuracy.

Another possibility is that grammatical predictability is moderated by linguistic proficiency; specifically, we will focus on Spanish proficiency, as most participants were stronger in English than in Spanish (bilinguals as well as language learners). Though bilinguals were on average more proficient than language learners in Spanish, there was significant overlap in proficiency between the groups (Figure 29). If grammatical predictability is moderated by linguistic proficiency, dividing the participants into two bilinguals and language learners might not show the effect.

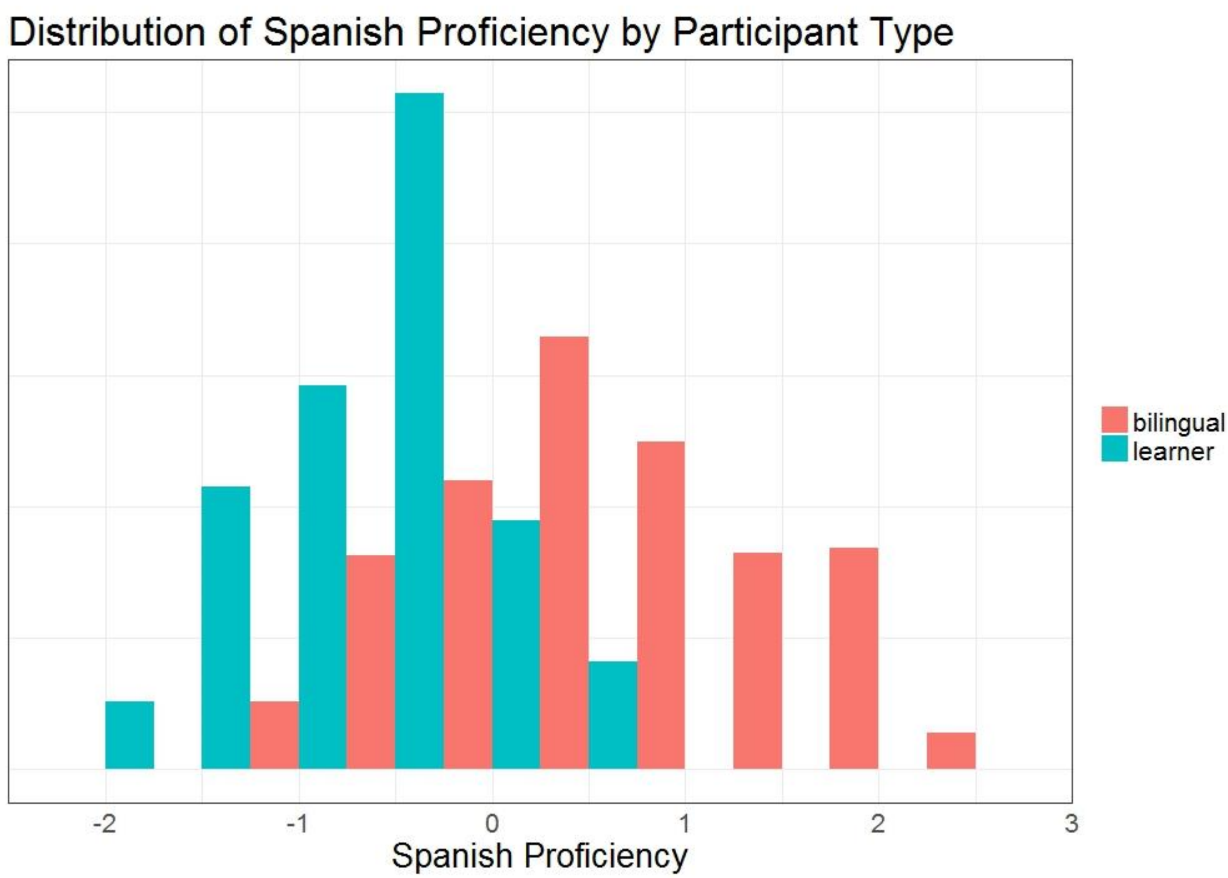


Figure 29: Distribution of Spanish proficiency by participant type.

We also wished to see if code-switching use would affect results. One possibility is that by code-switching in the past, bilinguals are processing code-switching sentences separately; we wished to see if grammatical predictability effects varied with code-switching use. We used participants' estimate of mean time switching with top conversational partners as our estimate.

We also looked at participants' judgments on ungrammatical code-switching sentences as a predictor for grammatical predictability. Language learners gave similar ratings for both grammatical and ungrammatical sentences, but bilinguals gave the grammatical sentences significantly higher ratings than they did the ungrammatical sentences. In addition when looking at ratings across the two groups, bilinguals and language learners had similar ratings for grammatical sentences ($t(98) = -1.15$, $p = 0.13$), but differed significantly on their judgments of ungrammatical sentences ($t(98) = -1.56$, $p = 0.006$), with bilinguals rating the sentences as more unacceptable than the language learners did. This suggests that higher sensitivity to Spanish-English code-switching grammatical constraints is best seen in participants' ratings of ungrammatical sentences. We wish to see if sensitivity to grammatical constraints affect grammatical predictability effects.

As before, we wish to control for differences in performance due to linguistic knowledge, so we include language proficiency measures in our models. Different measures of language proficiency were correlated in both bilinguals and language learners, so we again used the lexTale and lexTALE-ESP as our measures of proficiency in the model. This allows us to have equivalent measures for both English and Spanish that the participants in both Experiments 1 and 2 completed (bilinguals were not tested with the Spanish grammar test, and language learners were not tested with the English EIT).

Specific word measures are included; length, frequency, spillover length, spillover frequency, as well as trial number. All comparisons use first pass reading time of the second word as our main measure, and random effects for both item and subject are included in analyses. As in both Experiments 1 and 2, we expect to find a switch cost, and effects due to word properties. As both the language learners and bilinguals showed slower reading times for Spanish as compared to English, we also expect to find an effect of language.

Finally, we compared the relevant language learner and bilingual responses on the language proficiency measures. Regardless of whether or not linguistic variables affect grammatical predictability, the linguistic assessment questionnaires may capture some differences between the two groups. In particular, we were interested in seeing if there were differences among the two types of participant in regards to participants' self-ratings of English and Spanish, the mean amount of time spent code-switching, self-rating of time spent code-switching in the last week, mean code-switching positivity score, and the lexTALE, lexTALE-ESP, and code-switching judgment task scores.

Hypotheses

Question 1: Does language 'native-ness' affect grammatical predictability? Our null hypothesis (H_0) is that there will be no effects or interactions involving participant type, although there may be effects of proficiency that are correlated with participant type. According to this hypothesis, there is no impact of native-ness on grammatical predictability.

The alternative hypothesis (H_1) is that there are effects or interactions involving participant type. Bilinguals and language learners do not perform similarly in the experiment, and this is not due to an effect of native-ness.

Question 2: Does Spanish proficiency affect grammatical predictability? Our null hypothesis (H_0) is that Spanish proficiency does not affect grammatical predictability, at least not with our participants, since they are bilinguals and advanced language learners. According to this hypothesis, there is no impact of Spanish proficiency on grammatical predictability (less proficient participants still have enough linguistic knowledge to show and effect of grammatical predictability).

The alternative hypothesis (H_1) is that there are effects or interactions involving Spanish proficiency. Spanish proficiency moderates grammaticality. In particular, we would expect to find an effect of Spanish proficiency X grammaticality X language, as Spanish proficiency is likely to affect only

Spanish words, not English words. Similarly, an effect of Spanish proficiency x grammaticality x congruency x language would also indicate that Spanish proficiency is affecting grammatical predictability, though it may not be seen in all language conditions.

Question 3: Does code-switching usage affect grammatical predictability? Our null hypothesis (H_0) is that code-switching use does not affect grammatical predictability across languages. We will see no interaction between code-switching use and grammatical predictability, nor between code-switching use, grammatical predictability, and language congruency.

The alternative hypothesis (H_1) is that there is an interaction between code-switching use and grammatical predictability. Use of code-switching increases predictability during mixed language trials due to greater experience with code-switching in general.

Question 4: Does sensitivity to Spanish-English code-switching grammar constraints affect grammatical predictability? Our null hypothesis (H_0) is that sensitivity to Spanish-English code-switching grammar constraints does not affect grammatical predictability. We will see no interaction between ratings of ungrammatical code-switching sentences and grammatical predictability in mixed language trials.

The alternative hypothesis (H_1) is that there is an interaction between ratings of ungrammatical code-switching sentences, congruency, and grammatical predictability. Higher sensitivity to Spanish-English code-switching constraints indicates a greater linguistic knowledge in code-switching and this affects grammatical predictability.

Data

Data collected in Experiments 1 and 2.

Scoring

Language proficiency measures. The data from 50 bilingual participants and 50 language learner on the lexTALE, lexTALE-ESP, code-switching sentence judgment task, ACSES, and LHQ 2.0 proficiency tasks was analyzed. Specifically, we were interested in seeing if there were any differences among the two groups.

Lexical decision task. As before, we looked only at critical grammatical and ungrammatical trials. Filler trials were not analyzed, leaving a potential 14,150 trials to be analyzed. Incorrect trials were removed (3.89% of trials) leaving 13,591 total trials used in the reading time analyses.

As in Experiments 1 and 2, we used a participant's average percentage correct (APC) by language x congruency x grammaticality to conduct the accuracy analysis, leaving a total of eight values per participant. In total, there were 800 values used in the model, 8 per participant.

Results

Language Proficiency Measures

We looked at a total of 11 correlations and planned comparisons, so as to avoid false positives, a Bonferroni correction was applied; probability values have to be below $4.55*10^{-3}$ to be considered significant. Table 16 shows the average language proficiency and use scores for bilinguals and language learners.

Table 16: Language proficiency and use means for bilinguals and language learners.

	Spanish-English Bilinguals	Spanish L2 Learners
lexTALE score	90.25%	97.25%
English self-rating [1:7]	6.36	6.91

English EIT [1:4]	**3.71**	**NA**
lexTALE-ESP score	**79.92%**	**63.13%**
Spanish self-rating [1:7]	**6.20**	**6.25**
Spanish Grammar Test	**NA**	**68.31%**
Spanish EIT [1:4]	**3.78**	**2.79**
Estimated Percent Time Code-switching	**23.68%**	**21.46%**
Percent Time Code-switching with Top Conversational Partners	**35.73%**	**17.37%**
Code-switching Positivity Rating [1:7]	**4.23**	**4.82**
Grammatical Code-switching Sentence Rating	**5.29**	**5.66**
Ungrammatical Code-switching Sentence Rating	**3.63**	**4.40**

ACSES and LHQ 2.0. Language self-rating data followed expected patterns. The language learner mean self-rating for English was 6.91 (range: 6 : 7, sd = 0.21), and the bilingual mean self-rating for English was 6.36 (range: 3.5 : 7, sd = 0.78). The language learner self-rating for English was significantly higher than the bilingual self-rating for English, $t(98) = 4.81$, $p = 5.39*10^{-6}$. The language learner mean self-rating for Spanish was 4.93 (range: 3 : 6.25, sd = 0.73), and the bilingual mean self-rating for Spanish was 6.20 (range: 3.5 : 7, sd = 0.86). The language learner self-rating for Spanish was significantly lower than the bilingual self-rating for Spanish, $t(98) = -7.99$, $p = 2.63*10^{-12}$.

The language learner mean self-rating for the percentage of time spent code-switching was 21.46% (range: 0% : 80%, sd = 18.40%), and the bilingual mean self-rating for the percentage of time spent code-switching was 23.68% (range: 0% : 90%, sd = 22.71%). The two groups did not differ on their self-rating of time spent code-switching.

The language learner estimated percent mean time spent code-switching score with top conversational partners was 17.37% (range: 0% : 91.67%, sd = 18.57%), and the bilingual estimated percent mean time spent code-switching score with top conversational partners was 35.73% (range: 0% : 100%, sd = 24.76%). The language learner estimated percent mean time spent code-switching score with top conversational partners was significantly lower than the bilingual estimated percent mean time spent code-switching score with top conversational partners, $t(98) = -4.20$, $p = 5.98*10^{-5}$. This is likely reflective of the fact that bilinguals have a greater opportunity to code-switch than language learners.

The difference between the language learner mean self-rating for the percentage of time spent code-switching and the estimated percent mean time spent code-switching score with top conversational partners was 4.09% (range: -36.67% : 80%, sd = 23.54%). The difference between the bilingual mean self-rating for the percentage of time spent code-switching and the estimated percent mean time spent code-switching score with top conversational partners was -12.05% (range: -61.67% : 68.33%, sd = 22.93%). The difference between mean self-rating for the percentage of time spent code-switching and the estimated percent mean time spent code-switching score with top conversational partners, was significantly different between the two groups, $t(98) = 3.47$, $p = 7.65*10^{-4}$. Note: negative score for the difference indicates a higher value for mean time spent code-switching with top conversational partners. In other words, language learners estimate that they spent more time code-switching in the past week than they normally do with their top conversational partners. Bilinguals, on the other hand, say that when they talk with their top conversational partners they generally spend more time code-switching than they did in general the past week.

Language learners had a mean code-switching positivity score of 4.82 (range: 1 : 6.7, sd = 1.28), and bilinguals had a mean code-switching positivity score of 4.23 (range: 1 : 7, sd = 1.52). Code-switching positivity scores did not differ among language learners and bilinguals, $t(98) = -2.09$, $p = 0.04$. (With Bonferroni correction, *p*-values had to be below $4.55*10^{-3}$ to be significant.) Though not significant, language learners in general appear to have higher code-switching positivity scores than bilinguals, reflecting previous research showing that bilinguals are likely to show negativity towards code-switching.

lexTALE and lexTALE-ESP. The language learner mean lexTALE score was 97.25% (range: 85.00% : 100%, sd = 2.38%), and the bilingual mean lexTALE score was 90.25% (range: 61.25% : 100%, sd = 8.91%). The language learner mean lexTALE score was significantly higher than the bilingual mean lexTALE score, $t(98) = 5.37$, $p = 5.34*10^{-7}$. This is unsurprising, as all language learners were native speakers of English, unlike the bilinguals.

The language learner mean lexTALE-ESP score was 63.13% (range: 48.33% : 78.33%, sd = 7.17%), and the bilingual mean lexTALE-ESP score was 79.92% (range: 57.50% : 100%, sd = 10.98%). The language learner mean lexTALE-ESP score was significantly lower than the bilingual mean lexTALE-ESP score, $t(98) = -9.05$, $p = 1.40*10^{-14}$. The difference in scores is reflective of the fact that the bilinguals were significantly more skilled in Spanish the language learners.

Code-switching sentence ratings. The language learner average grammatical sentence score was 5.66 (range: 1 : 7, sd = 1.42), and the bilingual average grammatical sentence score was 5.29 (range: 1 : 7, sd = 1.83). The two groups did not show a difference between their scores, showing that they both found the sentences equally acceptable.

The language learner average ungrammatical sentence score was 4.40 (range: 1 : 6.6, sd = 1.35), and the bilingual average ungrammatical sentence score was 3.63 (range: 1 : 6.8, sd = 1.65). The two groups did not show a difference between their scores, though there was a tendency for language learners to find the ungrammatical sentences more acceptable than the bilinguals, $t(98) = -2.56$, $p = 0.01$. (With Bonferroni correction, p-values had to be below $4.55*10^{-3}$ to be significant.)

The difference between the language learner grammatical and ungrammatical sentence scores was 1.26 (range: -1.2 : 3.8, sd = 1.1). The difference between the bilingual grammatical and ungrammatical sentence scores was 1.66 (range: -1.0 : 5.0, sd = 1.3). (The difference indicates how much higher participants rated grammatical sentences than ungrammatical ones.) The two types of participant show similar differences in scores, indicating that neither bilinguals did not find the grammatical sentences significantly more acceptable than the ungrammatical sentences.

Question 1: Does Language 'Native-ness' Affect Grammatical Predictability?

Accuracy analysis. As in previous accuracy analyses, we included variables in the model which corresponded to our variables of interest: grammaticality, language congruency, and language of Word 2. Additionally, as we were interested as to whether or not there was a difference between language learners and bilinguals, we included participant type as a variable. All possible interactions including the four variables were included. Each binary and centered.

Proficiency as measured by the lexTALE and the lexTALE-ESP tests were included as fixed variables to take into account English and Spanish proficiency, respectively. Proficiency was centered before being added to the model.

As before, we used a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subject, a random intercept, and slopes for the language of the second word and language congruency (Figure 30).

```
Model <- glmer(APC ~
            EnglishProficiency +
            SpanishProficiency +
            ParticipantType*WordLanguage*LanguageCongruence*Grammaticality +
            (1 |subject) +
              (0 + LanguageCongruence |subject) +
              (0 + WordLanguage |subject) +
            data=allParticipants, REML = F)
```

Figure 30: Linear mixed-effects model for accuracy analysis including participant type.

Grammatical trials were answered more accurately than ungrammatical trials, β = -0.015, $t(779)$ = -4.631, $p < .001$, showing a main effect of grammatical predictability. Participant type was also found to be significant, β = 0.011, $t(779)$ = 1.989, p = 0.05, with bilinguals being on average more accurate than language learners (96.38% as compared to 95.71% across all trials). There was an interaction between participant type and grammaticality, β = -0.018, $t(779)$ = -2.746, p = 0.006, with language learners showing a stronger grammaticality effect than bilinguals (Figure 31). This is not surprising, as we had previously found an effect of grammaticality in the language learner analysis, but not the bilingual analysis.

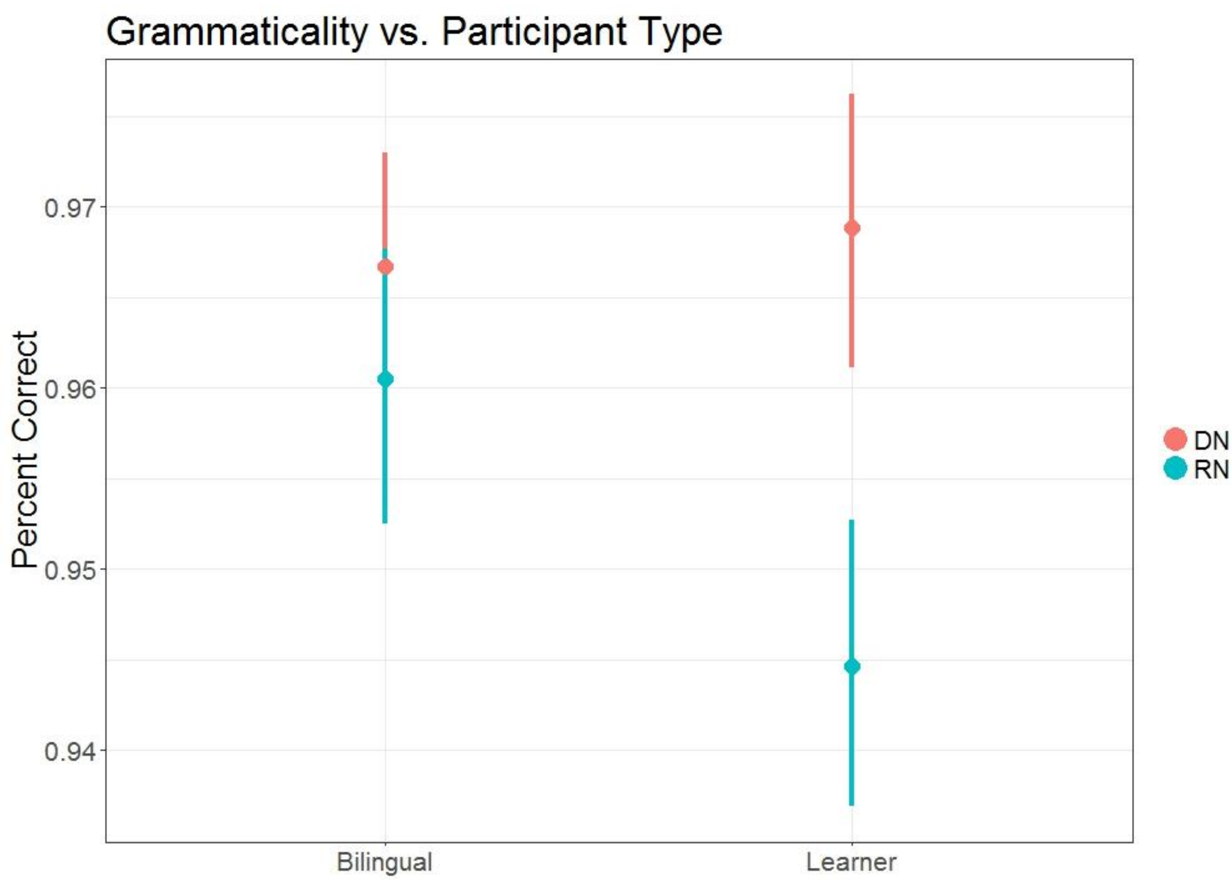


Figure 31: Grammaticality x participant type interaction in accuracy analysis including participant type. Error bars represent the standard error of the mean.

Same language trials were answered more accurately than mixed language trials, *β* = -0.011, $t(779) = -3.127$, $p = 0.002$, showing a language switch cost in mixed language trials. English trials were answered more accurately than Spanish trials, *β* = -0.023, $t(779) = -5.892$, $p < .001$, showing a main effect of language. There was also an interaction between congruency and the language of the second word, *β* = 0.028, $t(779) = 4.225$, $p < .001$, driven by a larger language switch cost when the second word was in English (decreased approximately 2.5%) as compared to when it was in Spanish (increased approximately 0.5%) (Figure 32). This is unsurprising, as we previously found similar interactions in both language learners and bilinguals. As before, we do not attribute this difference necessarily to asymmetrical language switch costs, as we would expect to see a language switch cost when the second word was in Spanish if this was the primary cause. Instead, we believe it is likely due to the fact that participants were either Spanish language learners, or bilinguals stronger in English than Spanish. Therefore, English-English trials were easier than Spanish-English trials, causing a drop in accuracy. Similarly, English-Spanish trials were easier than Spanish-Spanish trials, and so were more accurate.

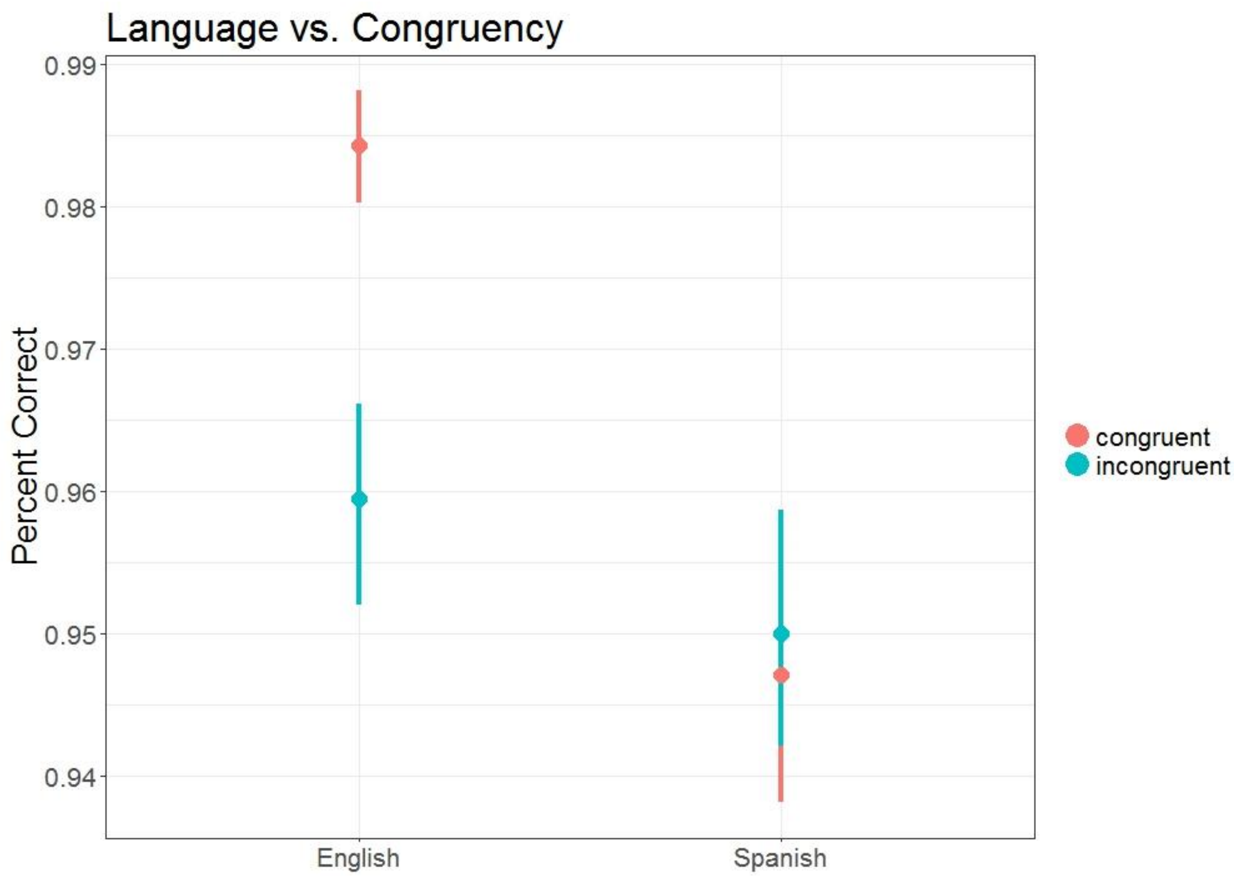


Figure 32: Congruency x language interaction in accuracy analysis including participant type. Language refers to the language of the second word; error bars represent the standard error of the mean.

There was an interaction between language and participant type, β = -0.034, $t(779)$ = -4.346, p < .001, with bilinguals showing similar accuracy in English and Spanish but language learners being more accurate in English than Spanish (Figure 33). This is not surprising, as we had before found an effect of language in the language learner analysis, but not the bilingual analysis.

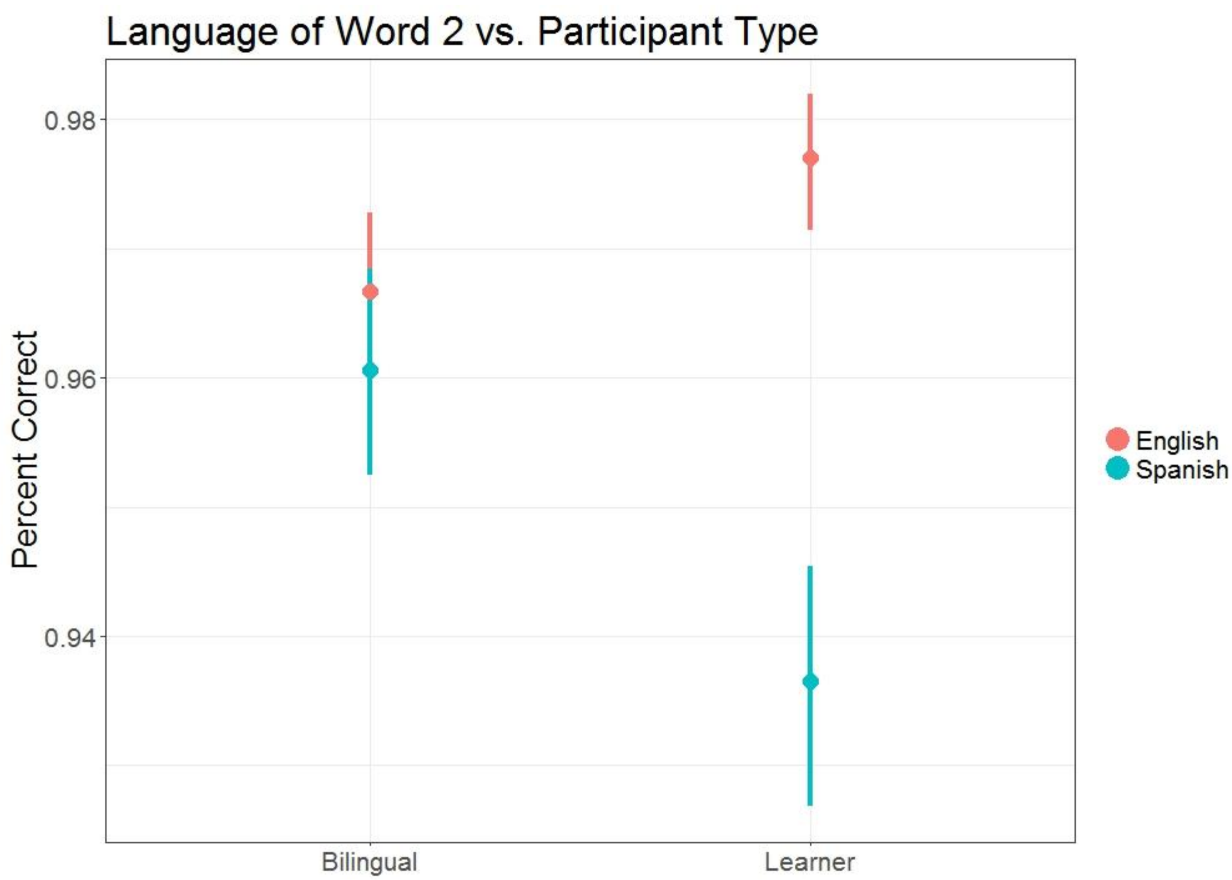


Figure 33: Language x participant type interaction in accuracy analysis including participant type. Language refers to the language of the second word; error bars represent the standard error of the mean.

There was also an interaction between congruency and participant type, *β* = 0.015, $t(779) = 2.121$, $p = 0.04$, with bilinguals showing a larger language switch cost than language learners (Figure 34). This is consistent with our previous analyses, as we had previously found an effect of language switch cost in the bilingual analysis, but not the language learner analysis.

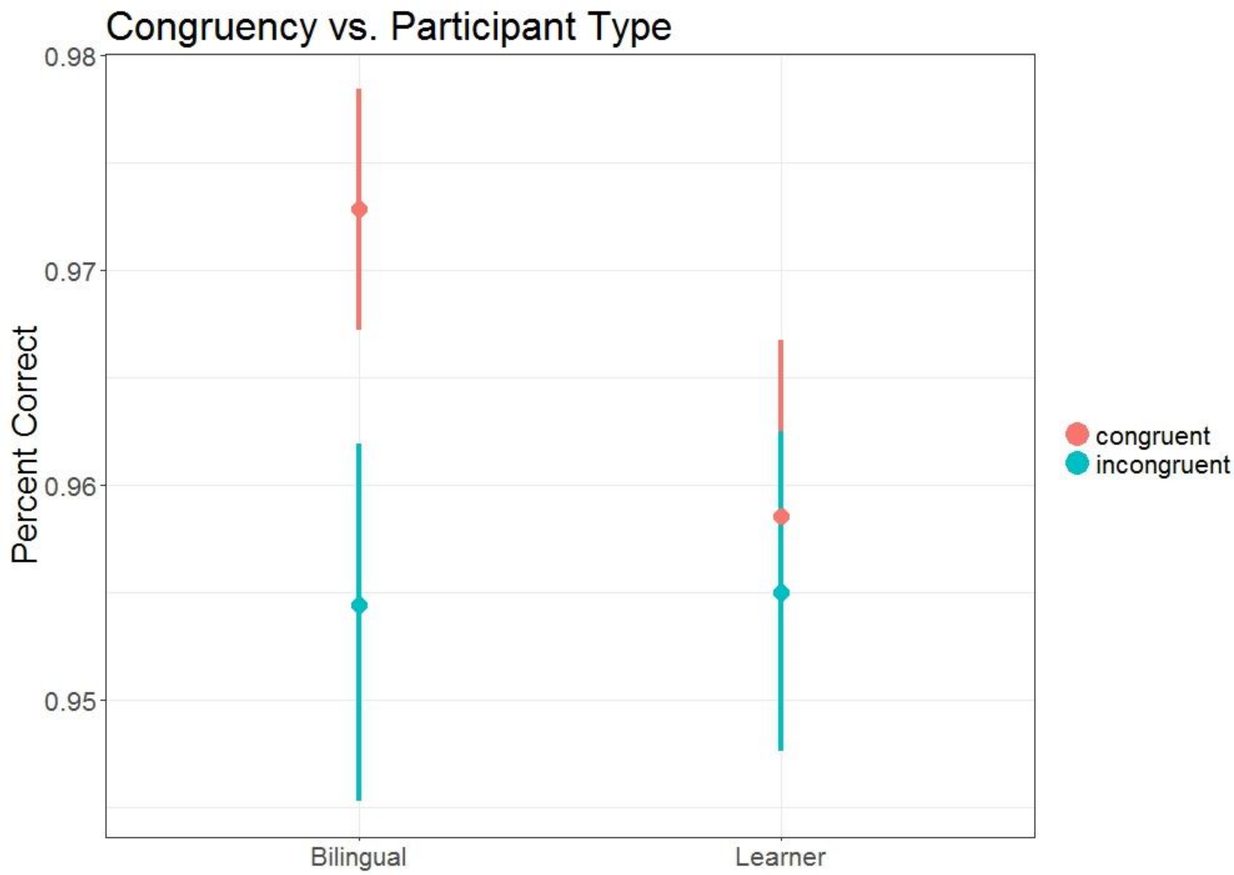


Figure 34: Congruency x participant type interaction in accuracy analysis including participant type. Error bars represent the standard error of the mean.

A three-way interaction was found between language, grammaticality, and participant type, β = 0.027, $t(779) = 2.067$, $p = 0.04$. As shown in Figure 35, this is because learners showed a larger grammaticality effect than bilinguals, especially when the critical word was in English.

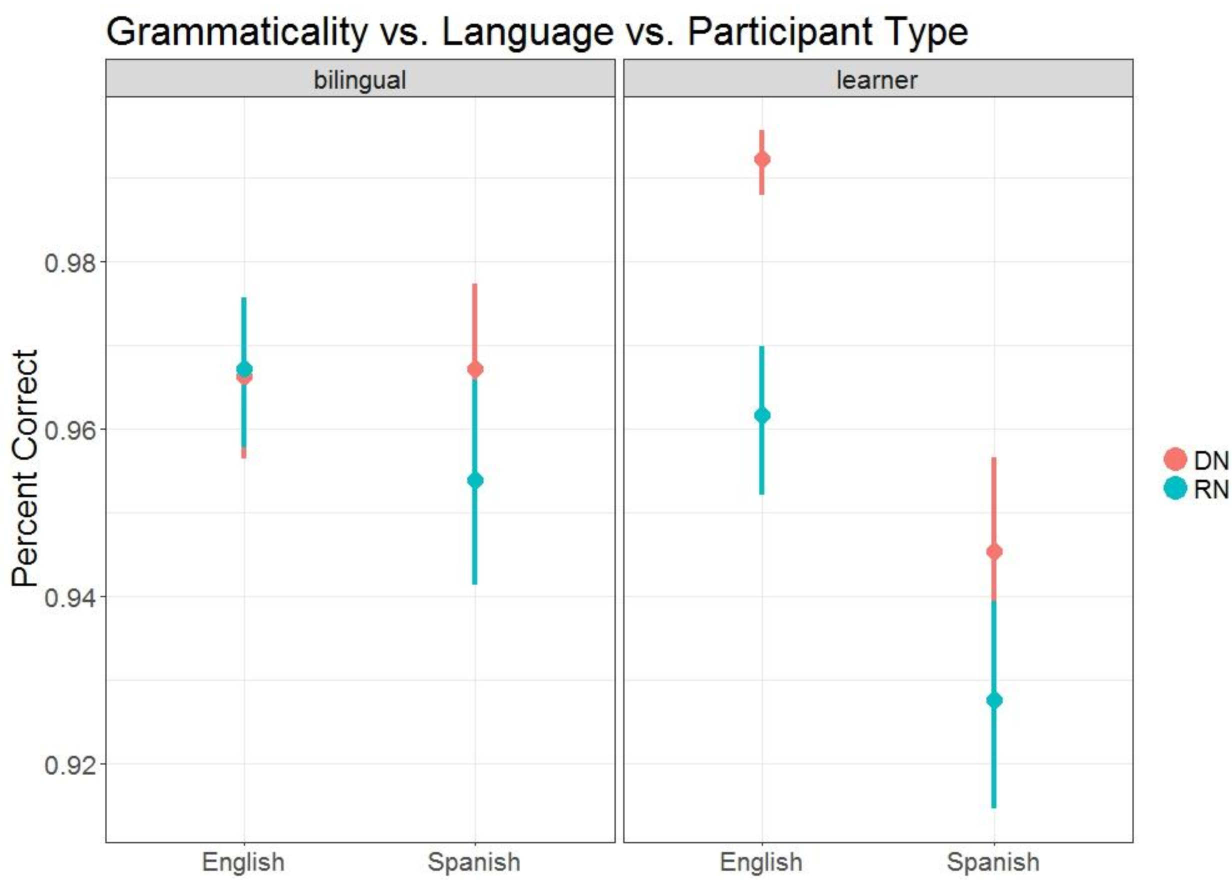


Figure 35: Language x grammaticality x participant type interaction in accuracy analysis including participant type.

In addition, a four-way interaction was found between language, congruency, grammaticality, and participant type, $\beta = 0.083$, $t(779) = 3.143$, $p = 0.002$. As shown in Figure 36, there are two potential sources of the interaction: either the reversal of the grammaticality effect for incongruent trials with the second word in English by bilinguals, or the absence of a grammaticality effect for congruent trials in Spanish by bilinguals.

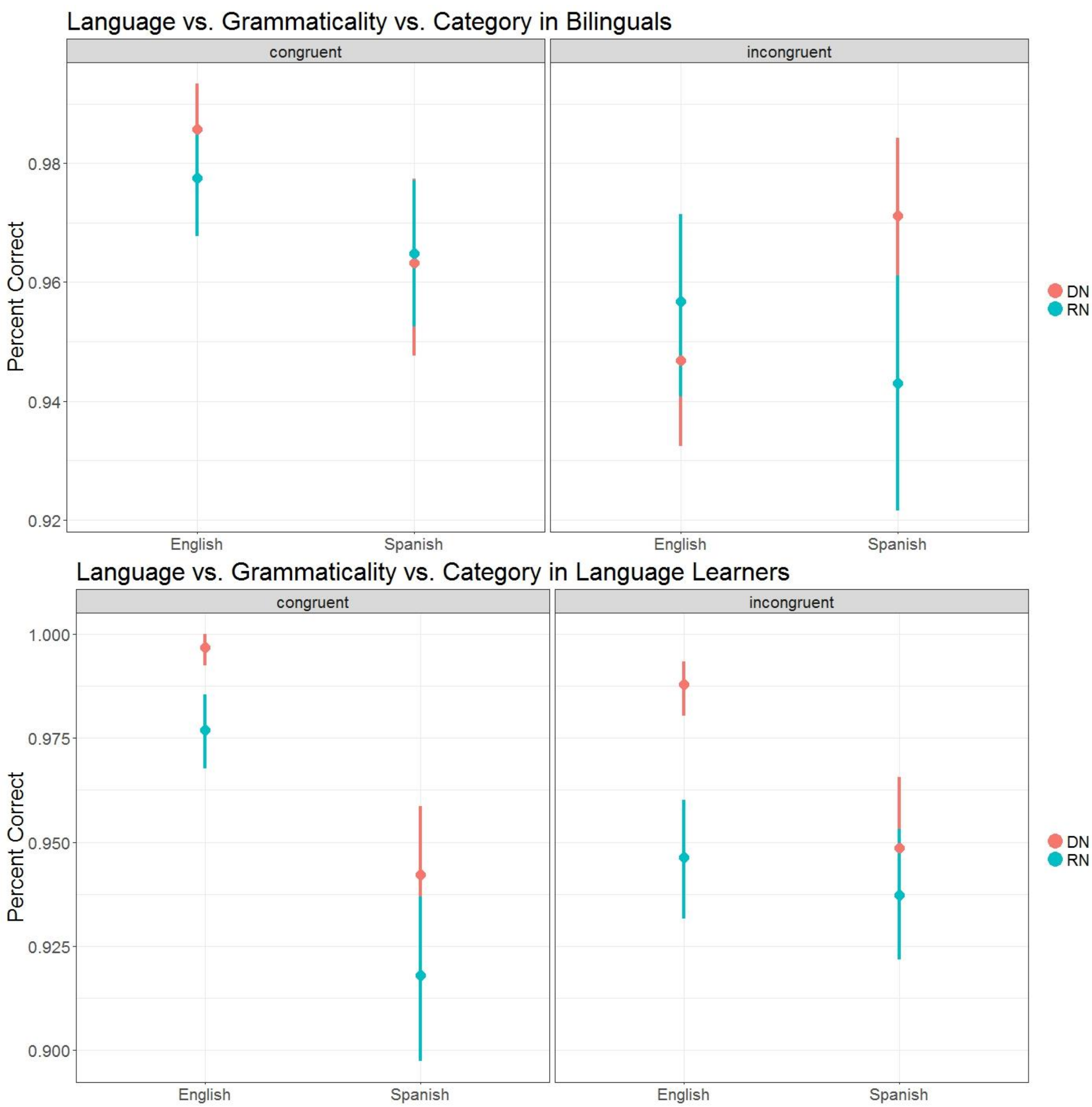


Figure 36: Language x congruency x grammaticality x participant type interaction in accuracy analysis including participant type. Error bars represent standard error of the mean.

Of the language proficiency variables, Spanish proficiency was found to be significant, $\beta = 0.015$, $t(779) = 5.797$, $p < .001$, with participants being generally more accurate as their proficiency increased. No effect of English proficiency was found.

Table 17 shows the results for the fixed variables of the model shown in Figure 30. *P*-values were generated using the lmerTest package (Kuznetsova et al., 2016).

Table 17: Results of the linear mixed-effects model for average accuracy including participant type. Significant results are bolded.

Variable	*β*	*t*-value	*p*-value
Intercept	**96.02%**	**504.245**	**<.001**
EnglishProficiency	0.003%	1.420	0.16
SpanishProficiency	**0.015%**	**5.797**	**<.001**
Language [Spanish]	**-0.023%**	**-5.892**	**<.001**
Congruency [Incongruent]	**-0.011%**	**-3.127**	**0.002**
Grammaticality [Ungrammatical]	**-0.015%**	**-4.631**	**<.001**
ParticipantType [Bilingual]	**0.011%**	**1.989**	**0.05**
Congruency*Grammaticality	-0.005%	-0.790	0.43
Congruency*Language	**0.028%**	**4.225**	**<.001**
Congruency*ParticipantType	**0.015%**	**2.121**	**0.04**
Grammaticality*Language	-0.001%	-0.093	0.93
Grammaticality*ParticipantType	**-0.018%**	**-2.746**	**0.006**
Language*ParticipantType	**-0.034%**	**-4.346**	**<.001**
Congruency*Grammaticality*Language	-0.006%	-0.495	0.62
Congruency*Grammaticality*ParticipantType	0.001%	0.107	0.92
Congruency*Language*ParticipantType	0.010%	0.742	0.46
Grammaticality*Language*ParticipantType	**0.027%**	**2.067**	**0.04**
Congruency*Grammaticality*Language*ParticipantType	**0.083%**	**3.143**	**0.002**

Table 18 shows accuracy rates by grammaticality, language condition, and congruency across all participants. For each row, the topmost number is the overall percentage correct per condition (not separated by grammaticality), followed by the percentage correct as divided by grammaticality (det-N is grammatical, adv-N is ungrammatical).

Table 18: Average accuracy rates across all participants. (See Table 4 for bilingual accuracy rates and Table 10 for language learner accuracy rates.)

	Percent Correct	
English-English	98.45%	
	det-N	99.11%
	adv-N	97.76%
Spanish-English	95.96%	
	det-N	96.72%
	adv-N	95.18%
English-Spanish	95.05%	
	det-N	96.00%
	adv-N	94.08%
Spanish-Spanish	94.73%	
	det-N	95.32%
	adv-N	94.12%
Same Language	96.59%	
	det-N	97.22%
	adv-N	95.94%
Mixed Language	95.51%	
	det-N	96.37%
	adv-N	94.63%
Word 2 (English)	97.20%	
	det-N	97.92%
	adv-N	96.47%
Word 2 (Spanish)	94.89%	
	det-N	95.66%
	adv-N	94.10%
ALL	96.05%	
	det-N	96.79%
	adv-N	95.29%

Reading time analysis. We fit a linear mixed-effects model to the first pass reading time of the second word of correctly answered 'yes' trials using the lme4 package in R (Bates et al., 2015; R Core

Team, 2015). The fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, English and Spanish proficiency scores as obtained on the lexTALE and the lexTALE-ESP tests respectively, and our main variables of interest: language of the second word, language congruency, grammaticality, and participant type, as well as all interactions of those four variables. All variables were centered before being added to the model.

Random effects were taken into account for both item and subject. As in Experiments 1 and 2, we followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for word length of the second word, frequency of the second word, word length of the first word, language of the second word, and trial number; for items, a random intercept, and slopes for English proficiency, language congruency, grammaticality, and trial number (Figure 37).

```
Model <- lmer( FPRT ~ TrialNumber  +
            WordLength +
            LogWordFrequency +
            LogSpilloverFrequency +
            SpilloverLength +
            EnglishProficiency +
            SpanishProficiency +
            ParticipantType*LanguageCongruence*Grammaticality*WordLanguage +
            (1 |subject) +
              (0 + WordLength |subject) +
              (0 + LogWordFrequency |subject) +
              (0 + SpilloverLength |subject) +
              (0 + WordLanguage |subject) +
              (0 + TrialNumber |subject) +
            (1 |item) +
              (0 + EnglishProficiency |item) +
              (0 + LanguageCongruence |item) +
              (0 + Grammaticality |item) +
              (0 + TrialNumber |item),
         data=allParticipants, REML=F)
```

Figure 37: Linear mixed-effects model for reading time analysis including participant type.

Of our main variables of interest, we found a grammaticality effect, $\beta = 10.382$, $t(13557) = 2.316$, $p = 0.02$, showing that nouns were read faster in a grammatical word pair as compared to an ungrammatical word pair. No interactions involving grammaticality were significant, but several had beta values of similar magnitude, which may indicate that they are affecting the results. In particular, the three-way interactions between congruency, grammaticality, and participant type ($\beta = 11.399$, $t(13557) = 1.244$, $p = 0.21$), language, grammaticality, and participant type ($\beta = 11.966$, $t(13557) = 1.264$, $p = 0.21$), as well as four-way interaction between congruency, language, grammaticality, and participant type ($\beta = 13.821$, $t(13557) = 0.757$, $p = 0.45$). These interactions suggest that bilinguals and language learners may not be showing the same behavior for all language conditions; not unexpected, as in Experiment 1 bilinguals showed a greater grammaticality effect in mixed language trials while in Experiment 2 learners showed similar grammaticality effects for both mixed and same language trials.

We also found a language effect, $\beta = 51.125$, $t(13557) = 7.681$, $p < .001$, with English words being read faster than Spanish words. (Expected, as half our participants are L2 Spanish speakers, and the bilinguals have been mostly educated in English.) We found a significant language congruency effect (language switch cost), $\beta = 19.713$, $t(13557) = 7.717$, $p < .001$, with words being read slower following a language switch than when following a word of the same language. No effect of participant type was found, suggesting that language learners and bilinguals are acting similarly. There was a trend between language of the second word and participant group, $\beta = -19.798$, $t(13557) = -1.829$, $p = 0.07$, with language learners showing a greater difference in the reading speed between English and Spanish words than bilinguals do. The language effect shows that English words were read faster than Spanish words regardless of participant type; this just refers that the difference was larger for language learners than for bilinguals. No other interactions between our main variables of interest were seen. Figure 38 shows the 95% confidence interval for the main variables of interest, in milliseconds; the dotted line indicates the baseline (language learner, English, congruent, grammatical).

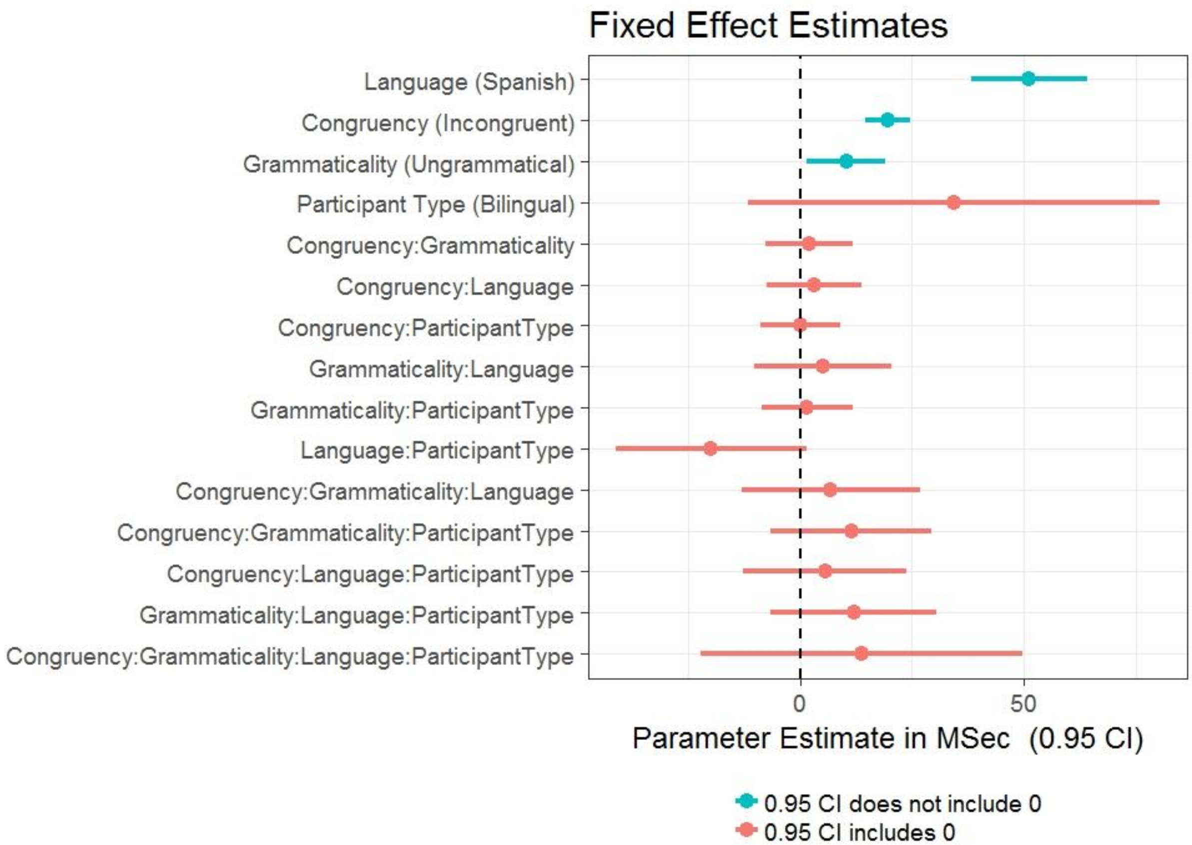


Figure 38: Fixed effects estimates for reading time analysis including participant type.

Of the control variables, we found main effects for the frequency and length of the second word, with more frequent and shorter words being read faster. (Frequency of Word 2: $\beta = -23.586$, $t(13557) = -7.442$, $p < .001$; Length of Word 2: $\beta = 53.721$, $t(13557) = 15.756$, $p < .001$) We found a significant Spanish proficiency effect ($\beta = -28.261$, $t(13557) = -2.457$, $p = 0.02$), showing that people with higher Spanish proficiency were faster. English proficiency failed to reach significance. There was a practice effect, with participants becoming faster as the experiment went on, $\beta = -13.042$, $t(13557) = -2.513$, $p = 0.01$. There was a trend of spillover frequency, $\beta = 4.083$, $t(13557) = 1.843$, $p = 0.07$, suggesting that nouns were read slower following words of higher frequency (not the expected direction). However, this is likely due to the specific stimuli used in the experiment, as nouns following adverbs were slightly more frequent than ones following determiners (mean log frequency of nouns following determiners: 2.57; mean log frequency of nouns following adverbs: 3.16).

Table 19 shows the results for the fixed variables of the model shown in Figure 37. *P*-values were generated using the lmerTest package (Kuznetsova et al., 2016).

Table 19: Results of the linear mixed-effects model for reading time including participant type. Significant results are bolded.

Variable	*β*	*t*-value	*p*-value
Intercept	**388.705**	**45.562**	**<.001**
TrialNumber	**-13.042**	**-2.513**	**0.01**
WordFrequency	**-23.586**	**-7.442**	**<.001**
WordLength	**53.721**	**15.756**	**<.001**
SpilloverFrequency	4.083	1.843	0.07
SpilloverLength	2.092	0.768	0.44
EnglishProficiency	-14.451	-1.496	0.14
SpanishProficiency	**-28.261**	**-2.457**	**0.02**
Language [Spanish]	**51.125**	**7.681**	**<.001**
Congruency [Incongruent]	**19.713**	**7.717**	**<.001**
Grammaticality [Ungrammatical]	**10.382**	**2.316**	**0.02**
ParticipantType [Bilingual]	34.310	1.466	0.15
Congruency*Grammaticality	2.106	0.424	0.67
Congruency*Language	3.142	0.582	0.56
Congruency*ParticipantType	0.169	0.037	0.97
Grammaticality*Language	5.131	0.659	0.51

Grammaticality*ParticipantType	1.559	0.301	0.76
Language*ParticipantType	-19.798	-1.829	0.07
Congruency*Grammaticality*Language	6.846	0.676	0.50
Congruency*Grammaticality*ParticipantType	11.399	1.244	0.21
Congruency*Language*ParticipantType	5.605	0.603	0.55
Grammaticality*Language*ParticipantType	11.966	1.264	0.21
Congruency*Grammaticality*Language*ParticipantType	13.821	0.757	0.45

Reading time analysis comparing grammaticality in same and mixed language trials. In the previous reading time analyses, we found a grammatical predictability effect and no effect of participant type. Though not significant, several interactions involving participant type and grammaticality (Table 19) suggest that bilinguals and language learners may not be showing the same pattern of grammatical predictability in same and mixed language trials. As in Experiments 1 and 2, we must compare the effect size of grammaticality for both same language and mixed language trials; furthermore, we must also include an interaction with participant type to see if both types of participants are behaving comparably. To do so, we separated grammaticality into two variables instead of looking for an interaction between grammaticality and congruency; grammaticality in congruent trials, and grammaticality in congruent trials, and included interactions with participant type.

We fit a linear mixed-effects model to the first pass reading time of the second word of correctly answered critical 'yes' trials using the lme4 package in R (Bates et al., 2015; R Core Team, 2015). As in the reading time analysis, the fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, English and Spanish proficiency scores as obtained on the lexTALE and the lexTALE-ESP tests respectively, and our main variables of interest: language of the second word, language congruency, grammaticality in congruent trials, grammaticality in incongruent trials, and participant type. Interactions between language and congruency, language and grammaticality in congruent trials, and grammaticality in incongruent trials were also included, as well as all possible interactions including participant type. All variables were centered before being added to the model.

Random effects were taken into account for both subject and item. Subject referred to the participant; item referred to the second word. Each participant was exposed to an item at most once in the experiment (participants were not exposed to all items, due to each word pair appearing in one of four possible language conditions, of which a participant was exposed to one). As in previous analyses, we followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for word length of the second word, word frequency of the second word, language of the second word, grammaticality in incongruent trials, and trial number; for items, a random intercept, and slopes for English proficiency, language congruency, grammaticality in congruent trials, grammaticality in incongruent trials, and trial number (Figure 39).

```
Model <- lmer(FPRT ~ TrialNumber  +
                  EnglishProficiency +
                  SpanishProficiency +
                  WordLength +
                  LogWordFrequency +
                  LogSpilloverFrequency +
                  SpilloverLength +
                  CongruentGrammaticality*WordLanguage*ParticipantType +
                  IncongruentGrammaticality*WordLanguage*ParticipantType +
                  LanguageCongruence*WordLanguage*ParticipantType +
                  (1 |subject) +
                    (0 + WordLength |subject) +
                    (0 + LogWordFrequency |subject) +
                    (0 + WordLanguage |subject) +
                    (0 + IncongruentGrammaticality |subject) +
                    (0 + TrialNumber |subject) +
                  (1 |item) +
                    (0 + EnglishProficiency |item) +
                    (0 + LanguageCongruence |item) +
                    (0 + CongruentGrammaticality |item) +
                    (0 + IncongruentGrammaticality |item) +
                    (0 + TrialNumber |item),
                  data=allParticipants, REML=F)
```

Figure 39: Linear mixed-effects model for reading time analysis including participant type and comparing grammaticality in same and mixed language trials.

We found an effect of grammatical predictability in incongruent trials, $\beta = 11.741$, $t(13561) = 2.065$, $p = 0.04$, with nouns being read faster in grammatical mixed language word pairs as compared to ungrammatical mixed language word pairs. We found a similar trend of grammatical predictability in congruent trials, $\beta = 8.977$, $t(13561) = 1.764$, $p = 0.08$, with nouns being read faster in grammatical same language word pairs as compared to ungrammatical same language word pairs, but it did not reach significance. Participant type was not significant, though its large beta value ($\beta = 34.202$, $t(13561) = 1.460$, $p = 0.15$) suggests that language learners may have shown overall faster reading speeds than bilinguals (though there was significant variability among participants).

There were no interactions between grammaticality and participant type, suggesting that the effect sizes for bilinguals and language learners was similar in both congruent and incongruent trials. One emergent pattern, though not significant, is that language learners show a greater grammatical predictability effect in same language trials ($\beta = -6.032$, $t(13561) = -0.921$, $p = 0.36$), and bilinguals show a greater grammatical predictability effect for mixed language trials ($\beta = 5.391$, $t(13561) = 0.691$, $p = 0.49$). In particular, bilinguals appear to show the greatest amount of grammatical predictability for English-Spanish trials, $\beta = 19.893$, $t(13561) = 1.508$, $p = 0.13$. This pattern may be in part the result of the trend for language learners to show a greater decrease in reading speed for Spanish words (in relation to English words) for all types of trials, $\beta = -20.303$, $t(13561) = 1.877$, $p = 0.06$.

We also found a language effect, $\beta = 51.154$, $t(13561) = 7.683$, $p < .001$, with Spanish words being read more slowly than English words. Language congruency was significant, $\beta = 19.750$, $t(13561) = 7.803$, $p < .001$, showing that nouns following a language switch were read slower than nouns following a word of the same language. No interaction was found between language and congruency, nor between participant type and congruency. Figure 40 shows the 95% confidence interval for the main variables of interest, in milliseconds; the dotted line indicates the baseline (English, congruent, grammatical, language learner).

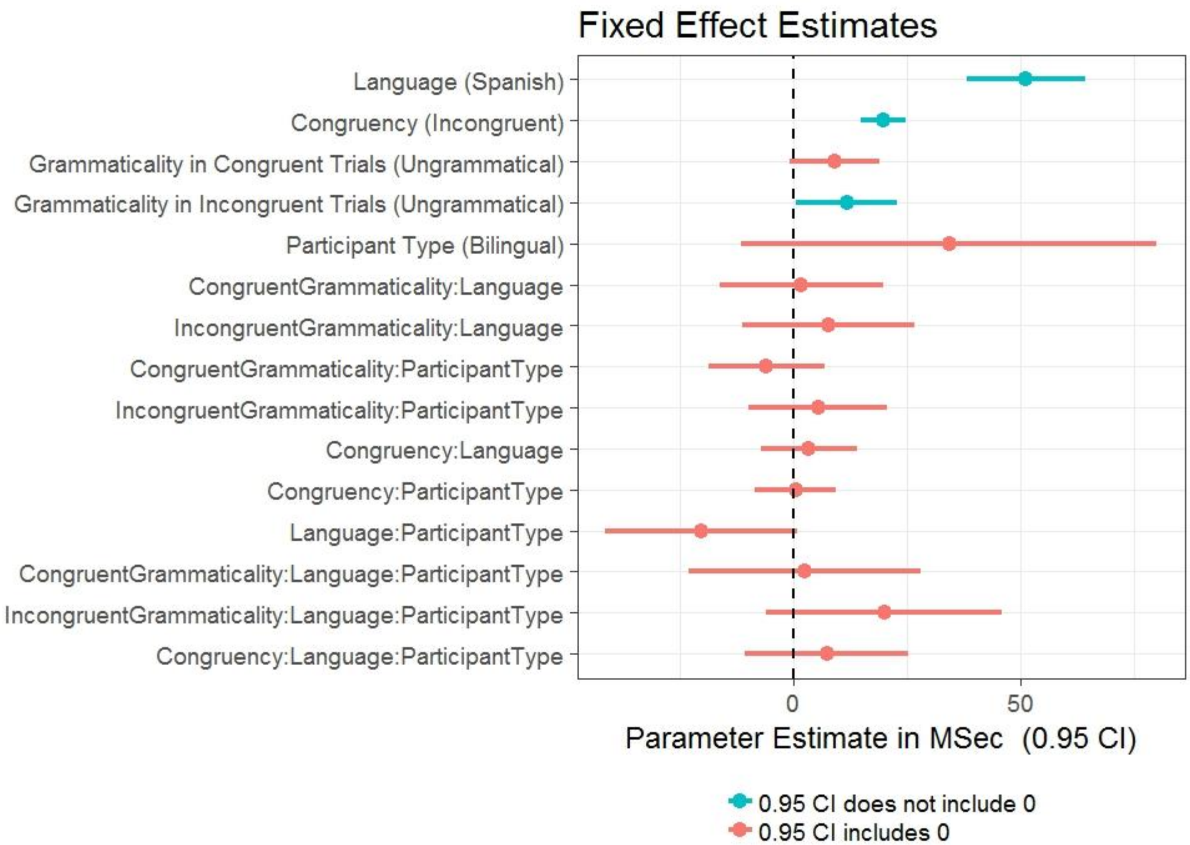


Figure 40: Fixed effect estimates for reading time analysis including participant type and comparing grammaticality in same and mixed language trials.

Of our control variables, we found a main effect for the frequency of the second word, with more frequent words being read faster than less frequent words, $\beta = -23.570$, $t(13561) = -7.452$, $p < .001$. There was also a main effect of for the length of the second word, with longer words being read slower than shorter words, $\beta = 53.622$, $t(13561) = 15.707$, $p < .001$). There was a practice effect, $\beta = -13.164$, $t(13561) = -2.534$, $p = 0.01$, with participants becoming faster as the experiment progressed. There was a trend for spillover frequency, $\beta = 4.045$, $t(13561) = 1.815$, $p = 0.07$, but in the opposite direction than expected, likely driven by the specific stimuli used in the experiment. Spillover length was not significant. There was an effect of Spanish proficiency, $\beta = -28.192$, $t(13561) = -2.449$, $p = 0.02$, with more proficient participants being faster; English proficiency was not significant.

Table 20 shows the results for the fixed variables of the model shown in Figure 39. *P*-values were generated using the lmerTest package (Kuznetsova et al., 2016).

Table 20: Results of the linear mixed-effects model for reading time including participant type and comparing grammaticality in same and mixed language trials. Significant results are bolded.

Variable	*β*	*t*-value	*p*-value
Intercept	**388.634**	**45.510**	**<.001**
TrialNumber	**-13.164**	**-2.534**	**0.01**
WordFrequency	**-23.570**	**-7.452**	**<.001**
WordLength	**53.622**	**15.707**	**<.001**
SpilloverFrequency	4.045	1.815	0.07
SpilloverLength	2.144	0.885	0.38
EnglishProficiency	-14.473	-1.497	0.14
SpanishProficiency	**-28.192**	**-2.449**	**0.02**
Language [Spanish]	**51.154**	**7.683**	**<.001**
Congruency [Incongruent]	**19.750**	**7.803**	**<.001**
CongruentGrammaticality [Ungrammatical]	8.977	1.764	0.08
IncongruentGrammaticality [Ungrammatical]	**11.741**	**2.065**	**0.04**
ParticipantType [Bilingual]	34.202	1.460	0.15
CongruentGrammaticality*Language	1.658	0.180	0.857
CongruentGrammaticality*ParticipantType	-6.032	-0.921	0.36
IncongruentGrammaticality*Language	7.581	0.785	0.43
IncongruentGrammaticality*ParticipantType	5.391	0.691	0.49
Congruency*Language	3.390	0.632	0.53

Congruency*ParticipantType	0.466	0.102	0.92
Language*ParticipantType	-20.303	-1.877	0.06
CongruentGrammaticality*Language*ParticipantType	2.425	0.185	0.85
IncongruentGrammaticality*Language*ParticipantType	19.893	1.508	0.13
Congruency*Language*ParticipantType	7.260	0.792	0.43

In Figure 41, we show the values of the predicted reading times as calculated using the coefficients shown in Table 20 for our main variables of interest, participant type, language congruence, and grammaticality. Grammatical predictability can be seen in all conditions, with words in grammatical contexts read faster than words in ungrammatical contexts, suggesting that both bilinguals and language learners form language-independent grammatical predictions. Notably, we see that bilinguals show greater grammatical predictability for nouns following a language change than language learners, which hints that bilinguals and learners may be behaving differently. In spite of this, it was not significant statistically, which may indicate that participant type is not the best predictor for grammatical predictability. As we saw in Figure 29, Spanish proficiency overlaps significantly for bilinguals and language learners; by separating participants into groups based on general language experience rather than proficiency, we may be adding unnecessary noise to the model.

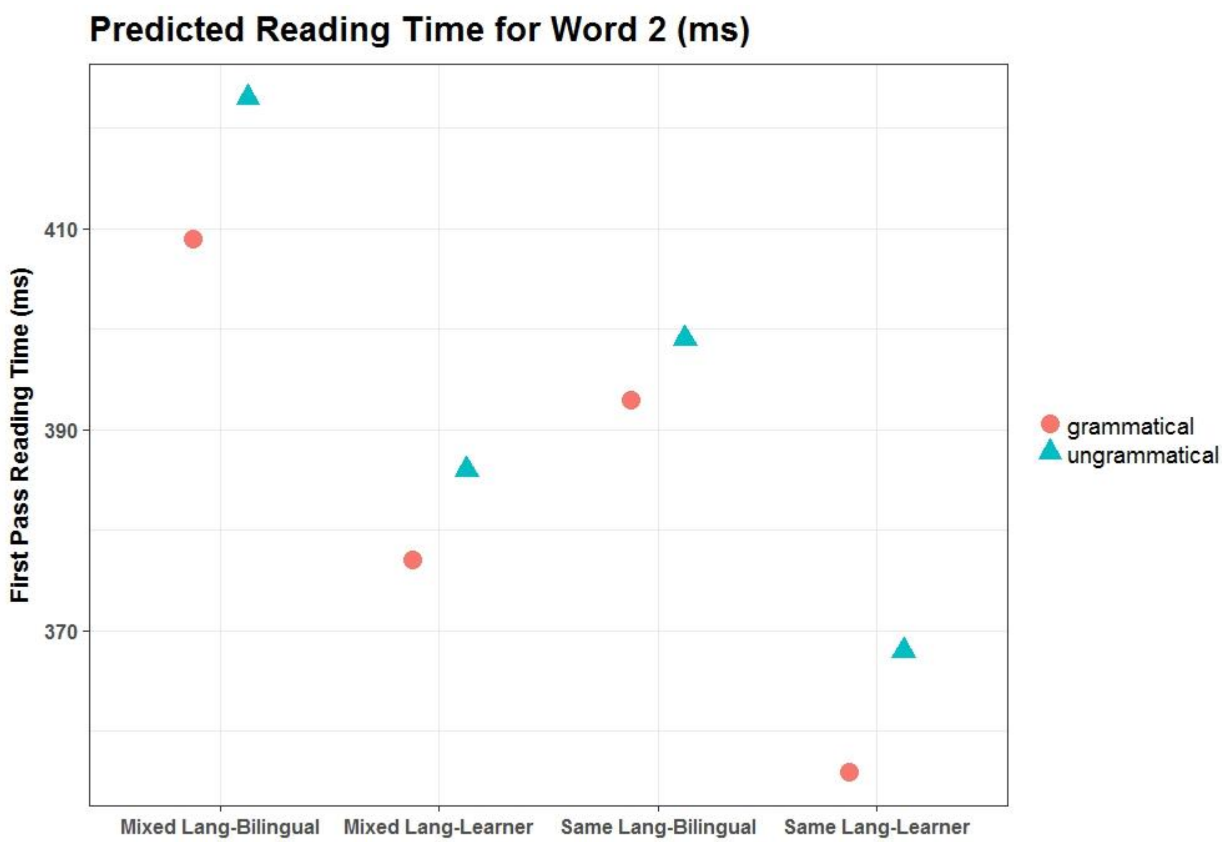


Figure 41: Point estimates for reading times given by linear mixed-effects model including participant type and comparing grammaticality in same and mixed language trials.

Question 2: Does Spanish Proficiency Affect Grammatical Predictability?

We fit a linear mixed-effects model to the first pass reading time of the second word of correctly answered 'yes' trials using the lme4 package in R (Bates et al., 2015; R Core Team, 2015). The fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, English proficiency score as obtained on the lexTALE test, and our main variables of interest: language of the second word, language congruency, grammaticality, and Spanish proficiency score as obtained on the lexTALE-ESP test. All possible interactions between main variables of interest were included. All variables were centered before being added to the model.

Random effects were taken into account for both item and subject. As before, we followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for word length of the second word, word log frequency of the second word, word length of the first word, language of the second word, and trial number; for items, a random intercept, and slopes for English proficiency, language congruency, grammaticality, and trial number (Figure 42).

```
Model <- lmer( FPRT ~ TrialNumber  +
            WordLength +
            LogWordFrequency +
            LogSpilloverFrequency +
            SpilloverLength +
            EnglishProficiency +
            SpanishProficiency*LangCongruence*Grammaticality*WordLanguage +
            (1 |subject) +
              (0 + LogWordFrequency |subject) +
              (0 + WordLength |subject) +
              (0 + SpilloverLength |subject) +
              (0 + WordLanguage |subject) +
              (0 + TrialNumber |subject) +
            (1 |item) +
              (0 + EnglishProficiency |item) +
              (0 + LangCongruence |item) +
              (0 + Grammaticality |item) +
              (0 + TrialNumber |item),
         data=allParticipants, REML=F)
```

Figure 42: Linear mixed-effects model for reading time analysis including interactions with Spanish proficiency.

We found a main effect of grammaticality, β = 10.389, t(13558) = 2.323, p = 0.02, where nouns were read faster in a grammatical word pair as compared to an ungrammatical word pair. There were no significant interactions with grammaticality. Though not significant, the interaction between language, congruency, and grammaticality had a beta value similar to that of the grammaticality effect, suggesting that not all language conditions are affected similarly by grammatical predictability (β = 7.138, t(13558) = 0.706, p = 0.48).

We found a main effect of language, β = 50.748, t(13558) = 8.187, p <.001, with English words were read faster than Spanish words. We also found a main effect of congruency, with nouns read slower following a language switch than when following a word of the same language, β = 19.775, t(13558) = 7.752, p <.001. There was a trend of Spanish proficiency, β = -18.327, t(13558) = -1.953, p = 0.06, with people of higher Spanish proficiency being faster per trial, though it did not reach significance.

There was an interaction between Spanish proficiency and language, β = -26.919, t(13558) = -5.487, p <.001. The interaction was due to Spanish (but not English) words being read faster as Spanish proficiency increased (Figure 43). No other interactions were found.

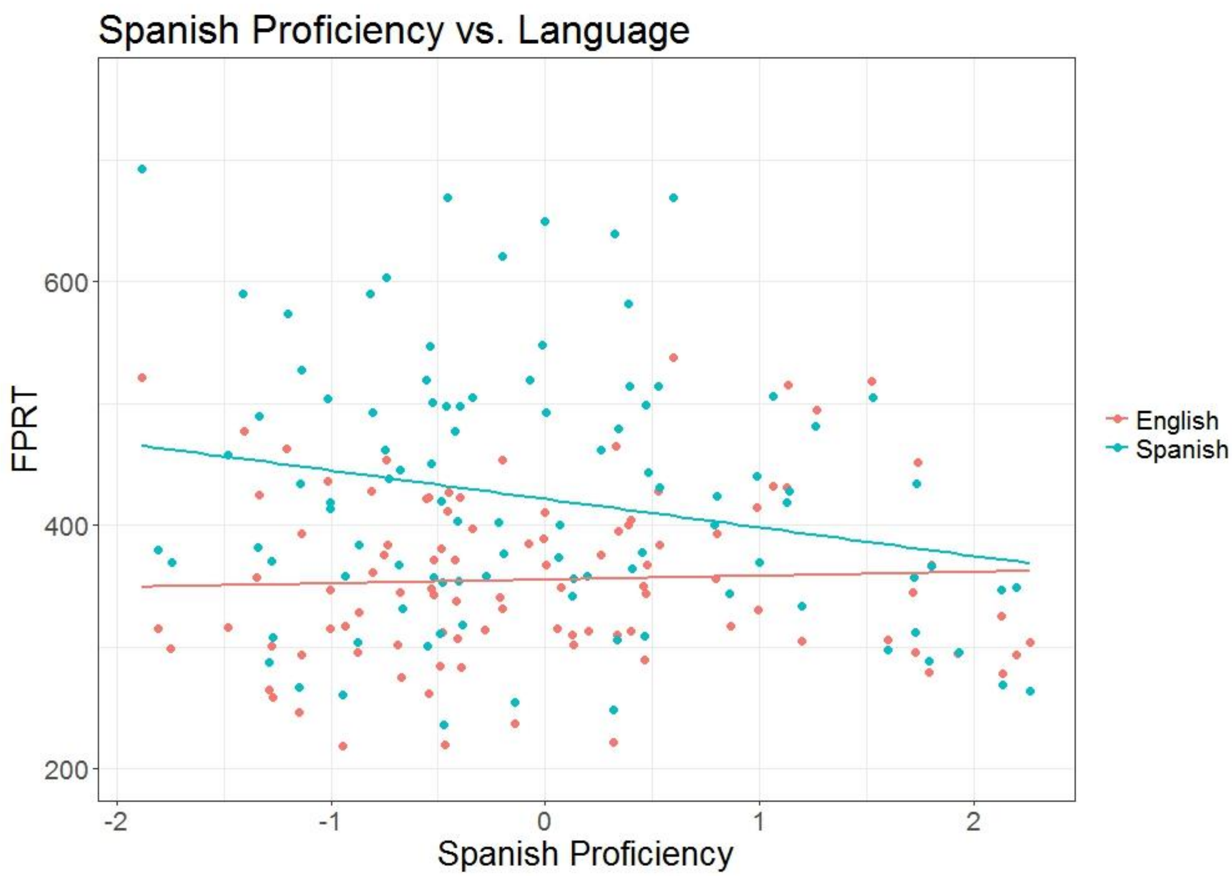


Figure 43: Language x Spanish proficiency interaction in reading time analysis including interactions with Spanish proficiency. Language refers to language of the second word.

Figure 44 shows the 95% confidence interval for the main variables of interest, in milliseconds; the dotted line indicates the baseline (English, congruent, grammatical).

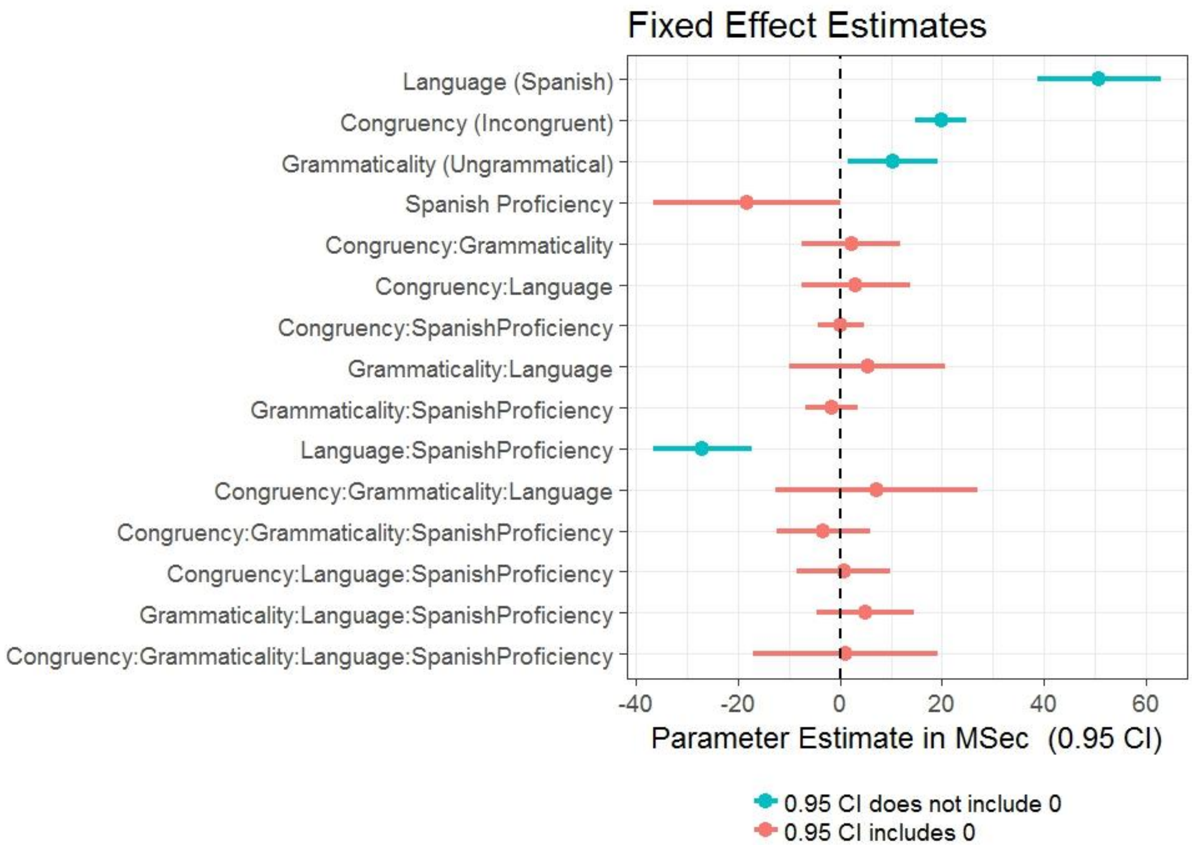


Figure 44: Fixed effects estimates for reading time analysis including interactions with Spanish proficiency.

We found main effects for the frequency and length of the second word, with words that are more frequent and words that are shorter being read faster. (Frequency of Word 2: $\beta = -23.553$, $t(13558) = -7.447$, $p < .001$; Length of Word 2: $\beta = 53.845$, $t(13558) = 15.802$, $p < .001$) A trend similar to that of Spanish proficiency was found for English proficiency, $\beta = -18.262$, $t(13558) = -1.943$, $p = 0.06$, just failing to reach significance. There was a practice effect, with participants becoming faster as the experiment went on, $\beta = -13.090$, $t(13558) = -2.526$, $p = 0.01$. As in previous models, there was a trend for spillover frequency words in the opposite direction, with higher frequency words being associated with slower reading times, $\beta = 4.044$, $t(13558) = 1.828$, $p = 0.07$; most likely due to the specific stimuli used in the experiment.

Table 21 shows the results for the fixed variables of the model shown in Figure 42. *P*-values were generated using the lmerTest package (Kuznetsova et al., 2016).

Table 21: Results of the linear mixed-effects model for reading time including interactions with Spanish proficiency. Significant results are bolded.

Variable	*β*	*t*-value	*p*-value
Intercept	**388.901**	**45.159**	**<.001**
TrialNumber	**-13.090**	**-2.526**	**0.01**
WordFrequency	**-23.553**	**-7.447**	**<.001**
WordLength	**53.845**	**15.802**	**<.001**
SpilloverFrequency	4.044	1.828	0.07
SpilloverLength	2.075	0.763	0.45
EnglishProficiency	-18.262	-1.943	0.06
SpanishProficiency	-18.327	-1.953	0.06
Language [Spanish]	**50.748**	**8.187**	**<.001**
Congruency [Incongruent]	**19.775**	**7.752**	**<.001**
Grammaticality [Ungrammatical]	**10.389**	**2.323**	**0.02**
Congruency*Grammaticality	2.210	0.446	0.66
Congruency*Language	3.132	0.581	0.56
Congruency*SpanishProficiency	0.095	0.041	0.97
Grammaticality*Language	5.378	0.693	0.49
Grammaticality*SpanishProficiency	-1.689	-0.637	0.52
Language*SpanishProficiency	**-26.919**	**-5.487**	**<.001**
Congruency*Grammaticality*Language	7.138	0.706	0.48

Congruency*Grammaticality*SpanishProficiency	-3.246	-0.701	0.48
Congruency*Language*SpanishProficiency	0.712	0.152	0.88
Grammaticality*Language*SpanishProficiency	4.918	1.015	0.31
Congruency*Grammaticality*Language*SpanishProficiency	1.154	0.125	0.90

Comparison to models without interactions with Spanish proficiency. As in other reading time models, we found main effects of grammaticality, language congruency, and language. Spanish proficiency did not reach significance, but there was a trend for faster reading times as it increased. As expected, Spanish proficiency interacted with language, with it affecting the reading speed of words in Spanish, but not English. Most important to our interest, Spanish proficiency did not interact with grammaticality, suggesting that it did not modulate grammatical predictability.

One difference between the model including interactions with Spanish proficiency and the models which instead had participant type as a variable is that there is less evidence that underlying factors may be affecting the pattern of results. The models including participant type suggested that bilinguals and language learners may be behaving differently, especially in English-Spanish trials. The model including interactions with Spanish proficiency showed less variation among the variables, though there were indications that same and mixed language trials may not be showing the same patterns of grammatical predictability. The model including interactions with participant type hinted that bilinguals may be showing greater grammatical predictability in mixed language trials; finding a similar pattern that higher Spanish proficiency leads to greater grammatical predictability would result in further evidence of that finding.

Reading time analysis including interactions with Spanish proficiency comparing grammaticality in same and mixed language trials. We fit a linear mixed-effects model to the first pass reading time of the second word of correctly answered critical ‘yes’ trials using the lme4 package in R (Bates et al., 2015; R Core Team, 2015). As in the reading time analysis, the fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, English and Spanish proficiency scores as obtained on the lexTALE and the lexTALE-ESP tests respectively, and our main variables of interest: language of the second word, language congruency, grammaticality in congruent trials, and grammaticality in incongruent trials. Interactions between language and congruency, language and grammaticality in congruent trials, and grammaticality in incongruent trials were also included, as well as all possible interactions including Spanish proficiency. All variables were centered before being added to the model.

Random effects were taken into account for both subject and item. Subject referred to the participant; item referred to the second word. Each participant was exposed to an item at most once in the experiment (participants were not exposed to all items, due to each word pair appearing in one of four possible language conditions, of which a participant was exposed to one). As in previous analyses, we followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for word length of the second word, word frequency of the second word, word length of the first word, language of the second word, grammaticality in incongruent trials, and trial number; for items, a random intercept, and slopes for English proficiency, language congruency, grammaticality in congruent trials, grammaticality in incongruent trials, and trial number (Figure 45).

```
Model <- lmer(FPRT ~ TrialNumber  +
                  EnglishProficiency +
                  WordLength +
                  LogWordFrequency +
                  LogSpilloverFrequency +
                  SpilloverLength +
                  CongruentGrammaticality*WordLanguage*SpanishProficiency +
                  IncongruentGrammaticality*WordLanguage*SpanishProficiency +
                  LanguageCongruence*WordLanguage*SpanishProficiency +
                  (1 |subject) +
                    (0 + WordLength |subject) +
                    (0 + LogWordFrequency |subject) +
                    (0 + SpilloverLength |subject)
                    (0 + WordLanguage |subject) +
                    (0 + IncongruentGrammaticality |subject) +
                    (0 + TrialNumber |subject) +
                  (1 |item) +
                    (0 + EnglishProficiency |item) +
                    (0 + LanguageCongruence |item) +
                    (0 + CongruentGrammaticality |item) +
                    (0 + IncongruentGrammaticality |item) +
                    (0 + TrialNumber |item),
                  data=allParticipants, REML=F)
```

Figure 45: Linear mixed-effects model for reading time analysis including interactions with Spanish proficiency comparing grammaticality in same and mixed language trials.

We found an effect of grammatical predictability in incongruent trials, β = 11.671, t(13556) = 2.093, p = 0.04, showing that nouns are being read faster in grammatical mixed language word pairs as compared to ungrammatical mixed language word pairs. We also found a similar trend of grammatical predictability in congruent trials, β = 9.064, t(13556) = 1.779, p = 0.08, with nouns being read faster in grammatical same language word pairs as compared to ungrammatical same language word pairs, though it did not reach significance.

There were no interactions between grammaticality in either congruent or incongruent trials and Spanish proficiency, suggesting that the effect sizes for bilinguals and language learners was similar in both types of trials. Similar to the model including interactions with participant type and looking at the effect of grammatical predictability in congruent and incongruent trials (Figure 39), English-Spanish trials seemed to show a greater grammatical predictability effect than Spanish-English trials, β = 8.050, t(13556) = 0.837, p = 0.40.

There was a trend for participants with greater Spanish proficiency to show faster reading times for nouns in the study, $\beta = -18.319$, $t(13556) = -1.951$, $p = 0.06$. Likewise, participants with greater English proficiency also showed a trend for faster reading times for nouns, $\beta = -18.252$, $t(13556) = -1.941$, $p = 0.06$. Unsurprisingly, higher Spanish proficiency affected the reading speed of words in Spanish, but not English, $\beta = -26.913$, $t(13556) = -5.486$, $p < .001$.

There was a language effect, $\beta = 50.858$, $t(13556) = 8.205$, $p < .001$, with Spanish words being read more slowly than English words. We also found an effect of language congruency, $\beta = 19.692$, $t(13556) = 7.871$, $p < .001$, showing that nouns following a language switch were read slower than nouns following a word of the same language. No interaction was found between language and congruency.

Figure 46 shows the 95% confidence interval for the main variables of interest, in milliseconds; the dotted line indicates the baseline (English, congruent, grammatical, language learner).

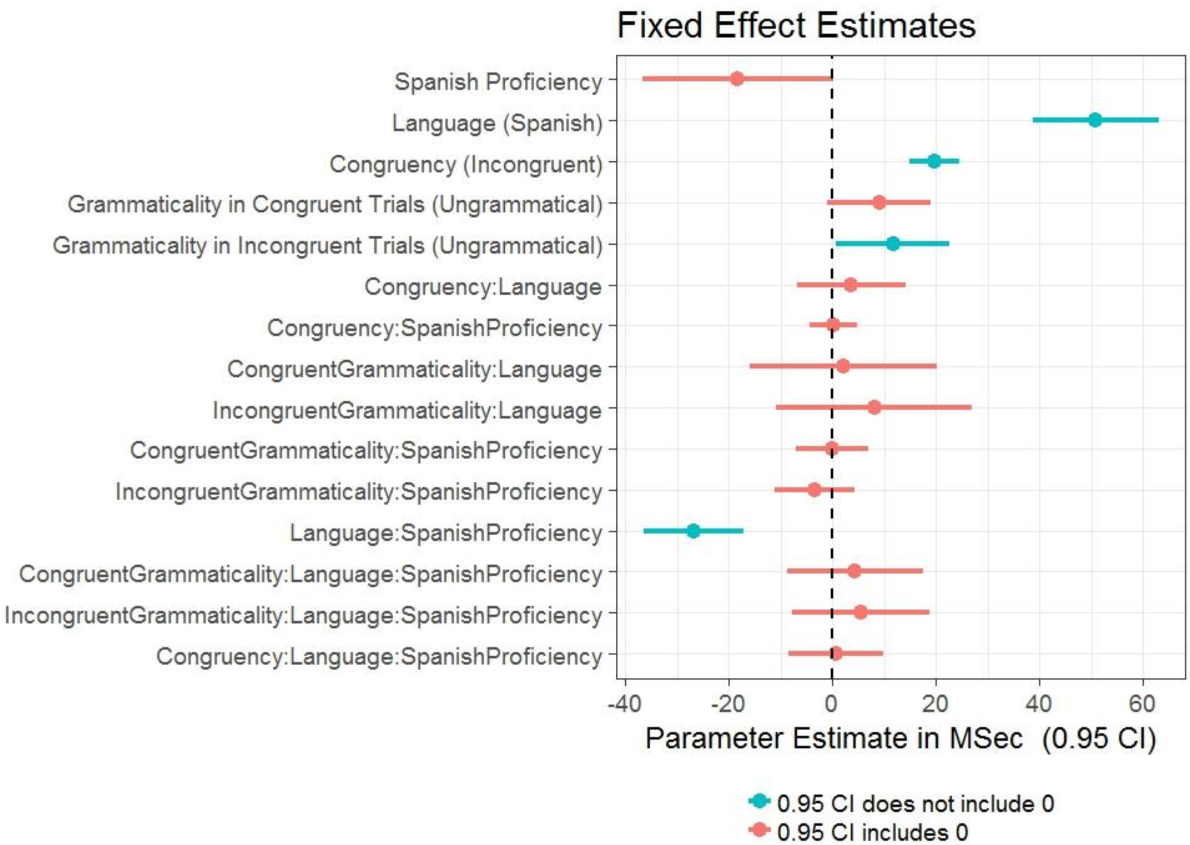


Figure 46: Fixed effect estimates for reading time analysis including interactions with Spanish proficiency comparing grammaticality in same and mixed language trials.

We found a main effect for the frequency of the second word, with more frequent words being read faster than less frequent words, β = -23.535, $t(13556)$ = -7.455, $p < .001$. There was also a main effect of for the length of the second word, with reading time increasing as the length of the word increased, β = 53.766, $t(13556)$ = 15.760, $p < .001$). Participants become faster as the experiment progressed, showing a practice effect, β = -13.158, $t(13556)$ = -2.541, p = 0.01. As in previous models, there was a trend for spillover frequency, β = 4.017, $t(13556)$ = 1.811, p = 0.07, but in the opposite direction than expected, likely driven by the specific stimuli used in the experiment. Spillover length was not significant.

Table 22 shows the results for the fixed variables of the model shown in Figure 45. *P*-values were generated using the lmerTest package (Kuznetsova et al., 2016).

Table 22: Results of the linear mixed-effects model for reading time including interactions with Spanish proficiency comparing grammaticality in same and mixed language trials. Significant results are bolded.

Variable	*β*	*t*-value	*p*-value
Intercept	**388.904**	**45.132**	**<.001**
TrialNumber	**-13.158**	**-2.541**	**0.01**
WordFrequency	**-23.535**	**-7.455**	**<.001**
WordLength	**53.766**	**15.760**	**<.001**
SpilloverFrequency	4.017	1.811	0.07
SpilloverLength	2.148	0.788	0.43
EnglishProficiency	-18.252	-1.941	0.06
SpanishProficiency	-18.319	-1.951	0.06
Language [Spanish]	**50.858**	**8.205**	**<.001**
Congruency [Incongruent]	**19.692**	**7.871**	**<.001**
CongruentGrammaticality [Ungrammatical]	9.064	1.779	0.08
IncongruentGrammaticality [Ungrammatical]	**11.671**	**2.093**	**0.04**
Congruency*Language	3.632	0.685	0.49
Congruency*SpanishProficiency	0.183	0.079	0.94
CongruentGrammaticality*Language	2.157	0.234	0.81
IncongruentGrammaticality*Language	8.050	0.837	0.40
CongruentGrammaticality*SpanishProficiency	-0.061	-0.017	0.99
IncongruentGrammaticality*SpanishProficiency	-3.505	-0.884	0.38
Language*SpanishProficiency	**-26.913**	**-5.486**	**<.001**
CongruentGrammaticality*Language*SpanishProficiency	4.372	0.656	0.51

IncongruentGrammaticality*Language*SpanishProficiency	5.436	0.809	0.42
Congruency*Language*SpanishProficiency	0.707	0.151	0.88

Comparisons to previous reading time models including all participants. We found a similar pattern of results when comparing the grammaticality predictability effect in congruent and incongruent models, and looking at how interactions with Spanish proficiency would affect the data (rather than participant type, as in Figure 39). Grammatical predictability was significant in incongruent trials, and a slightly smaller effect was found in congruent trials (though not significant). Interestingly, the grammaticality effect in incongruent trials was not sensitive to language, being greatest in English-Spanish trials, and did not interact with Spanish proficiency. This is in contrast to the model comparing including interactions with participant type comparing the grammaticality effect in congruent and incongruent trials, which suggested that the effect was driven by bilinguals. As bilinguals and language learners showed considerable overlap in Spanish proficiency, this suggests that participant type was not effectively capturing differences in behavior in relation to grammaticality.

Question 3: Does Code-switching Use Affect Grammatical Predictability?

We fit a linear mixed-effects model to the first pass reading time of the second word of correctly answered 'yes' trials using the lme4 package in R (Bates et al., 2015; R Core Team, 2015). The fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, English and Spanish proficiency scores as obtained on the lexTALE and the lexTALE-ESP tests respectively, and our main variables of interest: language of the second word, language congruency, grammaticality, and the mean amount of time spent code-switching with top conversational partners. All possible interactions between main variables of interest were included. All variables were centered before being added to the model.

Random effects were taken into account for both item and subject. As before, we followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for word length of the second word, word log frequency of the second word, language of the second word, and trial number; for items, a random intercept, and slopes for English proficiency, language congruency, and grammaticality (Figure 47).

```
Model <- lmer( FPRT ~ TrialNumber  +
            WordLength +
            LogWordFrequency +
            LogSpilloverFrequency +
            SpilloverLength +
            EnglishProficiency +
            SpanishProficiency +
            MeanTimeCS*LangCongruence*Grammaticality*WordLanguage +
            (1 |subject) +
              (0 + WordLength |subject) +
              (0 + LogWordFrequency |subject) +
              (0 + WordLanguage |subject) +
              (0 + TrialNumber |subject) +
            (1 |item) +
              (0 + EnglishProficiency |item) +
              (0 + LangCongruence |item) +
              (0 + Grammaticality |item),
         data=allParticipants, REML=F)
```

Figure 47: Linear mixed-effects model for reading time analysis including mean time code-switching with top conversational partners.

Nouns were read faster in a grammatical word pair as compared to an ungrammatical word pair, showing a grammatical predictability effect, $\beta = 10.771$, $t(13559) = 2.382$, $p = 0.02$. Grammaticality did not interact with language or congruency, though similar to previous models, the beta value for three-way interaction between language, congruency, and grammaticality was greater than might be expected ($\beta = 7.213$, $t(13559) = 0.710$, $p = 0.48$), reflecting that English-Spanish trials seemed to show a greater grammatical predictability effect than other language conditions.

English words were read faster than Spanish words, showing a language effect, $\beta = 51.622$, $t(13559) = 7.711$, $p < .001$. Nouns were read slower following a language switch than when following a word of the same language, showing a language congruency effect (language switch cost), *β* = 19.755, *t*(13559) = 7.692, *p* <.001. We did not find a main effect of mean time spent code-switching with top conversational partners, suggesting that amount of time spent regularly code-switching was not affecting the results. No interactions were found.

Figure 48 shows the 95% confidence interval for the main variables of interest, in milliseconds; the dotted line indicates the baseline (English, congruent, grammatical).

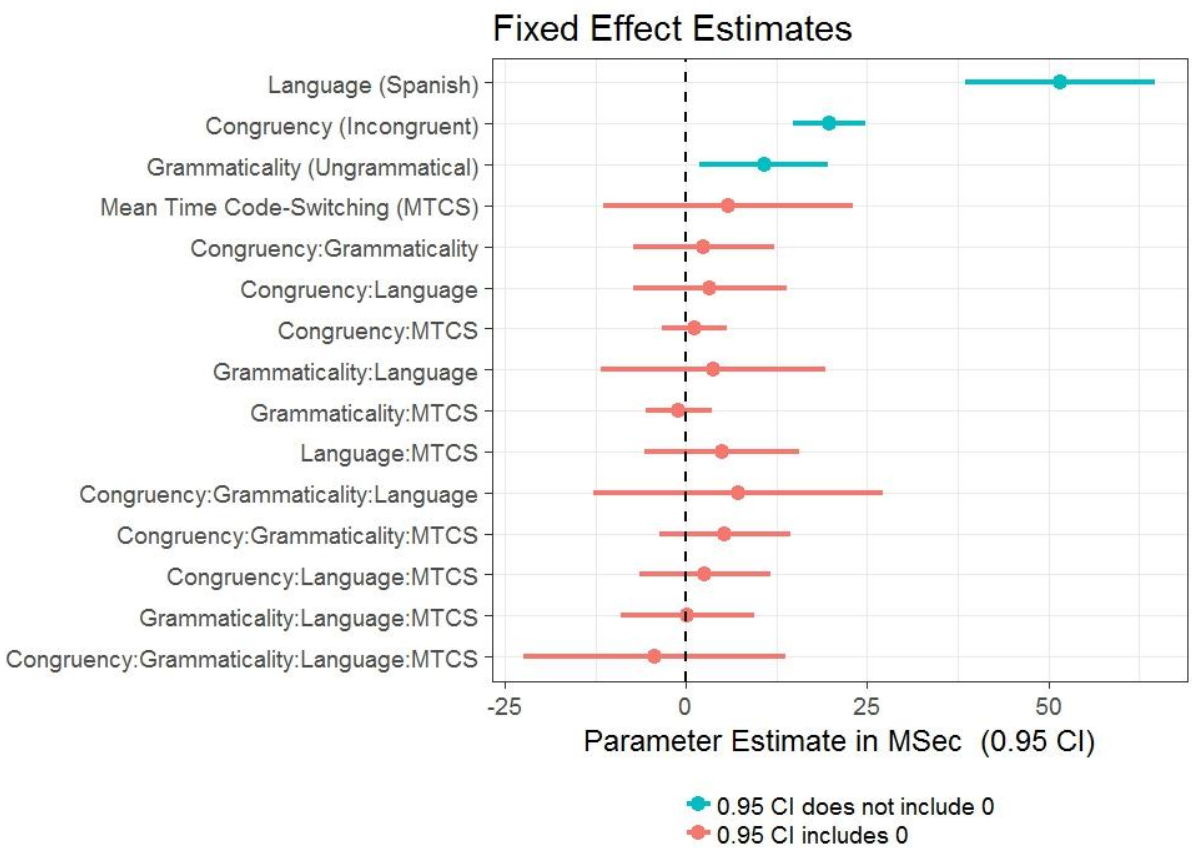


Figure 48: Fixed effects estimates for reading time analysis including mean time code-switching with top conversation partners.

As in previous models, we found main effects for the frequency and length of the second word, with words that are more frequent and words that are shorter being read faster. (Frequency of Word 2: *β* = -23.972, *t*(13559) = -7.549, *p* < .001; Length of Word 2: *β* = 53.313, *t*(13559) = 15.643, *p* < .001) Though it did not reach significance, there was a trend for participants with higher Spanish proficiency to be faster per trial, *β* = -18.035, *t*(13559) = -1.925, *p* = 0.06. A similar trend was found for English proficiency, *β* = -16.462, *t*(13559) = -1.688, *p* = 0.09. There was a practice effect, with participants becoming faster as the experiment went on, *β* = -13.675, *t*(13559) = -2.743, *p* = 0.007. As in previous models, there was a trend for spillover frequency words in the opposite direction, with higher frequency words being associated with slower reading times, *β* = 3.820, *t*(13559) = 1.728, *p* = 0.08; however, this is likely due to the specific stimuli used in the experiment.

Table 23 shows the results for the fixed variables of the model shown in Figure 47. *P*-values were generated using the lmerTest package (Kuznetsova et al., 2016).

Table 23: Results of the linear mixed-effects model for reading time including mean time spent code-switching with top conversational partners. Significant results are bolded.

Variable	*β*	*t*-value	*p*-value
Intercept	**389.542**	**45.302**	**<.001**
TrialNumber	**-13.675**	**-2.743**	**0.007**
WordFrequency	**-23.972**	**-7.549**	**<.001**
WordLength	**53.313**	**15.643**	**<.001**
SpilloverFrequency	3.820	1.728	0.08
SpilloverLength	2.151	0.906	0.36
EnglishProficiency	-16.462	-1.688	0.09
SpanishProficiency	-18.035	-1.925	0.06
Language [Spanish]	**51.622**	**7.711**	**<.001**
Congruency [Incongruent]	**19.755**	**7.692**	**<.001**
Grammaticality [Ungrammatical]	**10.771**	**2.382**	**0.02**

MeanTimeCS	5.789	0.659	0.51
Congruency*Grammaticality	2.453	0.493	0.62
Congruency*Language	3.294	0.607	0.54
Congruency*MeanTimeCS	1.178	0.512	0.61
Grammaticality*Language	3.790	0.478	0.63
Grammaticality*MeanTimeCS	-1.014	-0.435	0.66
Language*MeanTimeCS	4.903	0.900	0.37
Congruency*Grammaticality*Language	7.213	0.710	0.48
Congruency*Grammaticality*MeanTimeCS	5.333	1.157	0.25
Congruency*Language*MeanTimeCS	2.624	0.568	0.57
Grammaticality*Language*MeanTimeCS	0.191	0.041	0.97
Congruency*Grammaticality*Language*MeanTimeCS	-4.317	-0.469	0.64

Comparison to models without mean time code-switching as a variable. Mean time code-switching does not affect the outcome of the model greatly; grammaticality, language, and congruency were significant, and they did not interact with other variables. Most important in regards to our question, mean time code-switching with top conversational partners was not significant, nor were any interactions including it as a variable.

Question 4: Does Sensitivity to Spanish-English Code-switching Grammar Constraints Affect Grammatical Predictability?

We fit a linear mixed-effects model to the first pass reading time of the second word of correctly answered 'yes' trials using the lme4 package in R (Bates et al., 2015; R Core Team, 2015). The fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, English and Spanish proficiency scores as obtained on the lexTALE and the lexTALE-ESP tests respectively, and our main variables of interest: language of the second word, language congruency, grammaticality, and the average rating given to the ungrammatical code-switching sentences on the code-switching sentence judgment task. All possible interactions between main variables of interest were also included. All variables were centered before being added to the model.

Random effects were taken into account for both item and subject. As before, we followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for word length of the second word, word log frequency of the second word, language of the second word, and trial number; for items, a random intercept, and slopes for English proficiency, language congruency, grammaticality, and trial number (Figure 49).

```
Model <- lmer( FPRT ~ TrialNumber  +
            WordLength +
            LogWordFrequency +
            LogSpilloverFrequency +
            SpilloverLength +
            EnglishProficiency +
            SpanishProficiency +
            UCS*LangCongruence*Grammaticality*WordLanguage +
            (1 |subject) +
              (0 + WordLength |subject) +
              (0 + LogWordFrequency |subject) +
              (0 + WordLanguage |subject) +
              (0 + TrialNumber |subject) +
            (1 |item) +
              (0 + EnglishProficiency |item) +
              (0 + LangCongruence |item) +
              (0 + Grammaticality |item) +
              (0 + TrialNumber |item),
         data=allParticipants, REML=F)
```

Figure 49: Linear mixed-effects model for reading time analysis including average ungrammatical code-switching sentence rating.

We found a grammatical predictability effect, $\beta = 10.366$, $t(13558) = 2.307$, $p = 0.02$, showing that nouns were read faster in a grammatical word pair as compared to an ungrammatical word pair. As in previous models, we found a larger than expected beta value for the three-way interaction between language, congruency, and grammaticality (β = 7.176, t(13558) = 0.706, p = 0.48), indicating that the grammaticality effect is most prominent in English-Spanish trials.

We also found a language effect, β = 51.064, t(13558) = 7.586, p < .001, with English words being read faster than Spanish words. We found a significant language congruency effect (language switch cost), β = 19.802, t(13558) = 7.708, p < .001, with words being read slower following a language switch than when following a word of the same language. A main effect of average ungrammatical code-switching sentence rating was not found (β = -5.338, t(13558) = -0.611, p = 0.54), suggesting that sensitivity to code-switching constraints was not affecting overall reading times.

An interaction between ungrammatical code-switching sentence rating and language congruency was found, β = 6.480, t(13558) = 2.799, p = 0.005 (Figure 50). The pattern suggests that the participants who found the ungrammatical sentences least acceptable also showed the least language switch cost. A trend was also found between ungrammatical code-switching sentence rating, language congruency, and grammaticality, β = -8.909, t(13558) = -1.920, p = 0.06, suggesting that the interaction is driven more by grammatical word pairs rather than the ungrammatical word pairs. The pattern of lowered switch cost in participants who rated the ungrammatical code-switching sentences as less acceptable is present regardless of the grammaticality of the word pairs, however. The four-way interaction between ungrammatical code-switching sentence rating, language, language congruency, and grammaticality also had a relatively high beta value, β = -13.045, t(13558) = -1.411, p = 0.16, though it was not significant.

These interactions may suggest that ungrammatical sentence rating is not necessarily capturing sensitivity to Spanish-English constraints only; participants may be rating sentences as less acceptable due to other factors. For example, code-switching has often been seen as a sign of lack of linguistic knowledge in bilingual community, as being undesirable (Gumperz, 1977); of the 25 participants who rated the code-switching sentences lowest, 18 were bilingual. Though no interaction between language switch cost and participant type was found in the model testing for group differences (Figure 37), there may be differences among the bilingual participants which are being ignored by grouping them only as 'bilingual'.

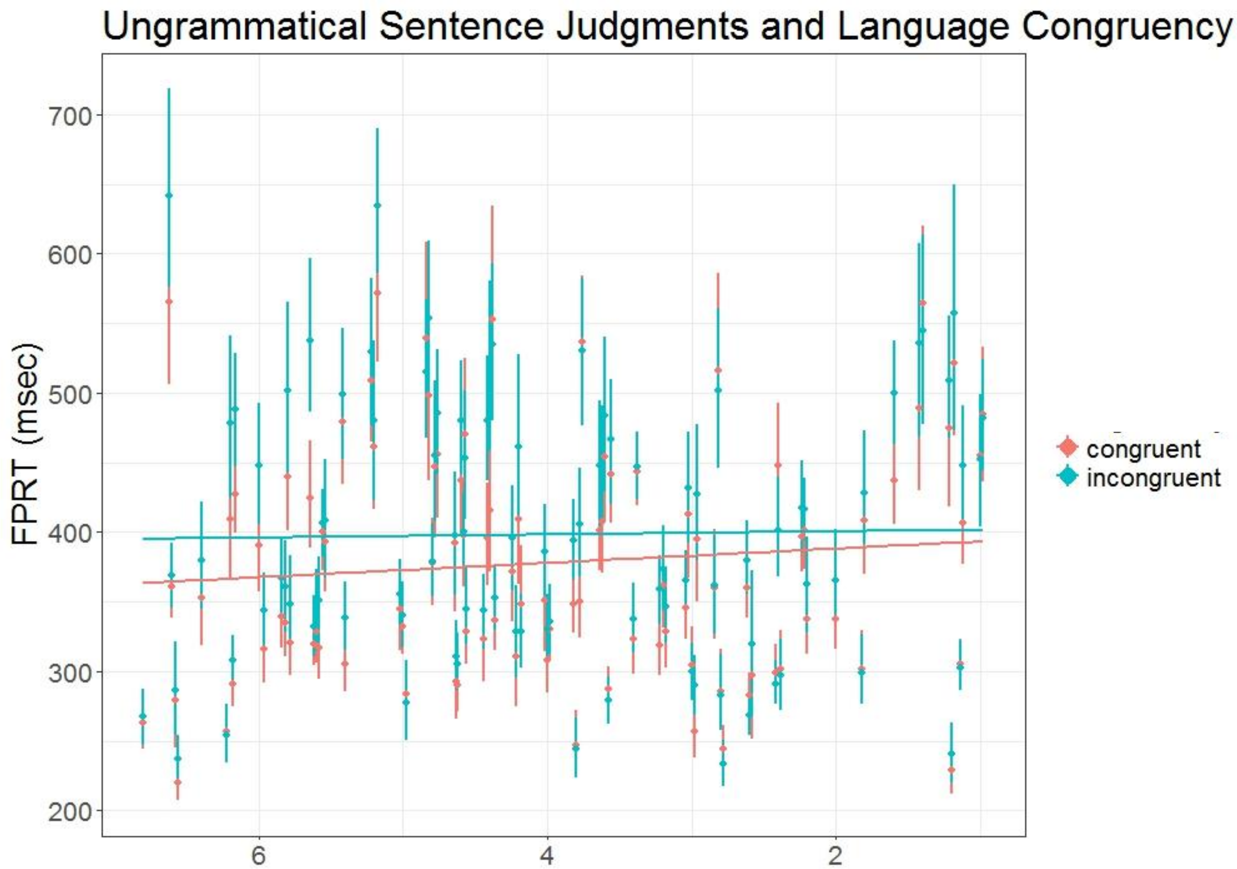


Figure 50: Congruency x ungrammatical code-switching sentence rating interaction in reading time analysis including average ungrammatical code-switching sentence rating. Lower value means that participant found the sentence less acceptable, so going to the right means decreasing acceptability of ungrammatical code-switching sentences.

No other interactions between our main variables of interest were seen. Figure 51 shows the 95% confidence interval for the main variables of interest, in milliseconds; the dotted line indicates the baseline (English, congruent, grammatical).

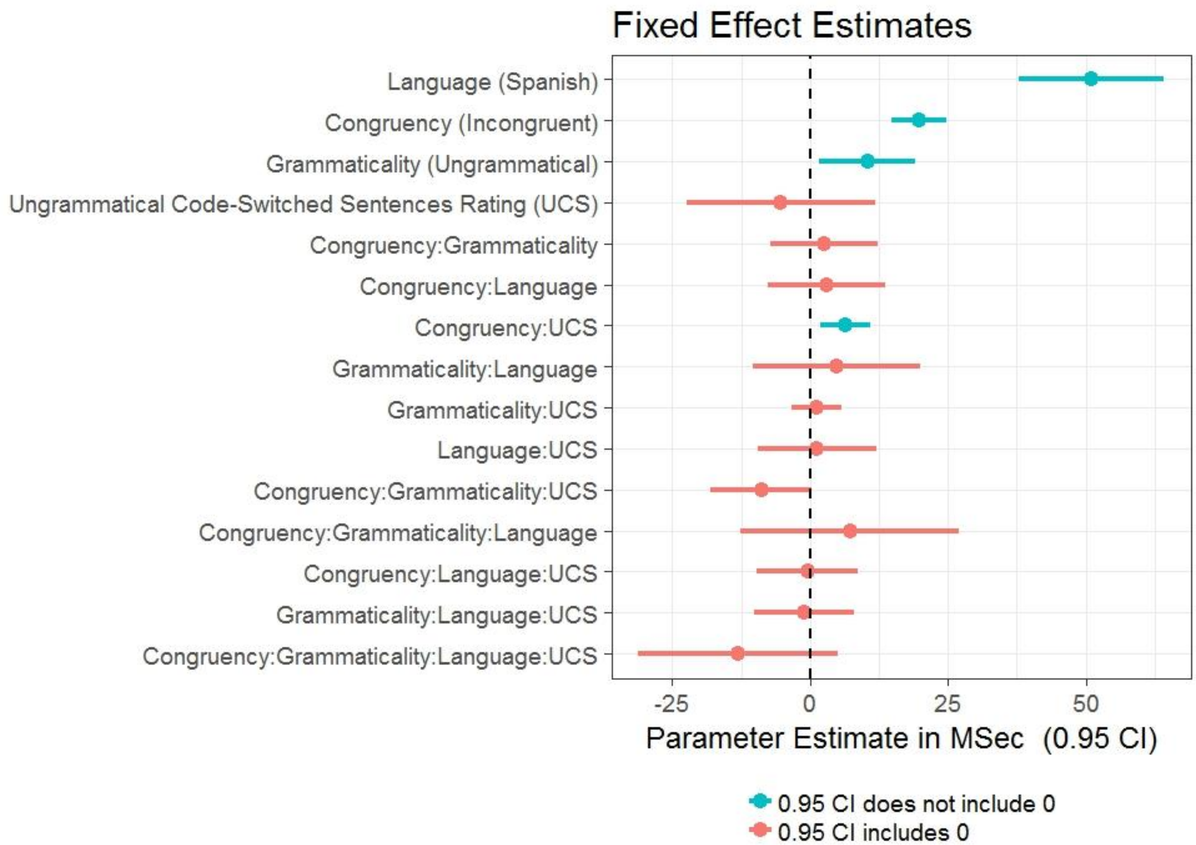


Figure 51: Fixed effects estimates for reading time analysis including average ungrammatical code-switching sentence rating.

As in previous models, we found main effects for the frequency and length of the second word, with more frequent and shorter words being read faster. (Frequency of Word 2: β = -23.631, $t(13558)$ = -7.453, p < .001; Length of Word 2: β = 53.670, $t(13558)$ = 15.729, p < .001) We found a significant Spanish proficiency effect, β = -19.891, $t(13558)$ = -2.037, p = 0.04, showing that people with higher Spanish proficiency were faster per trial. There was a similar effect of English proficiency, β = -18.880, $t(13558)$ = -1.999, p = 0.05, with people with higher English proficiency being faster per trial. There was a practice effect, with participants becoming faster as the experiment went on, β = -13.133, $t(13558)$ = -2.528, p = 0.01. In regards to spillover effects, there was a trend for higher frequency words to lead to slower reading times, *β* = -3.996, *t*(13558) = 1.798, *p* = 0.07, but it was likely due to the specific stimuli used in the experiment.

Table 24 shows the results for the fixed variables of the model shown in Figure 49. *P*-values were generated using the lmerTest package (Kuznetsova et al., 2016).

Table 24: Results of the linear mixed-effects model for reading time including average ungrammatical code-switching sentence rating. Significant results are bolded.

Variable	*β*	*t*-value	*p*-value
Intercept	**388.795**	**45.181**	**<.001**
TrialNumber	**-13.133**	**-2.528**	**0.01**
WordFrequency	**-23.631**	**-7.453**	**<.001**
WordLength	**53.670**	**15.729**	**<.001**
SpilloverFrequency	3.996	1.798	0.07
SpilloverLength	2.024	0.839	0.40
EnglishProficiency	**-18.880**	**-1.999**	**0.05**
SpanishProficiency	**-19.891**	**-2.037**	**0.04**
Language [Spanish]	**51.064**	**7.586**	**<.001**
Congruency [Incongruent]	**19.802**	**7.708**	**<.001**
Grammaticality [Ungrammatical]	**10.366**	**2.307**	**0.02**
UngrammaticalSentenceAverageRating (UCS)	-5.338	-0.611	0.54
Congruency*Grammaticality	2.538	0.510	0.61
Congruency*Language	3.080	0.567	0.57
Congruency*UCS	**6.480**	**2.799**	**0.005**
Grammaticality*Language	4.788	0.615	0.54

Grammaticality*UCS	1.092	0.471	0.64
Language*UCS	1.244	0.228	0.82
Congruency*Grammaticality*Language	7.176	0.706	0.48
Congruency*Grammaticality*UCS	-8.909	-1.920	0.06
Congruency*Language*UCS	-0.524	-0.113	0.91
Grammaticality*Language*UCS	-1.164	-0.251	0.80
Congruency*Grammaticality*Language*UCS	-13.045	-1.411	0.16

Comparison to models without ungrammatical code-switching sentence rating as a variable. Most importantly, models predict similar patterns, showing a main of grammaticality that does not interact with any other predictor. In regards to our question, ungrammatical code-switching sentence rating was not significant in the model including it as a predictor, though an interaction between and language congruency was significant, suggesting that it may be capturing some differences between participants (if not in the processing of grammatical vs. ungrammatical language pairs).

Discussion

Several questions motivated analyses combining the data from the first two experiments. First, we wished to see if language learners and bilinguals behaved the same way. There were several reasons why we wished to do so. First, because the reading time data was not entirely clear for the language learners. Looking at the bilinguals, we found a clear language-independent grammatical predictability effect; if anything, the effect of grammatical predictability was greatest in mixed language trials. Looking at the language learners, we found only a trend for grammatical predictability, not allowing us to draw strong conclusions. Furthermore, language learners showed grammatical predictability in same language trials, not mixed language trials. We also found a trend an interaction between language and grammaticality; however, the trend suggested that language learners showed a stronger grammatical predictability effect in Spanish, not English, opposite to what we might predict if language learners show language-specific grammatical predictability.

By combining the data, we are able to see if there are interactions with participant type, which would suggest that bilinguals and language learners are behaving differently; if not, it suggests that they are behaving the same. We are also able to see if participant type interacts with grammatical predictability in same or mixed language trials specifically, seeing if bilinguals and language learners show similar grammatical predictability regardless of the language of the word pair.

Other than wanting to find further evidence for grammatical predictability in language learners, we wished also to see if 'native-ness' affected grammatical predictability. The accuracy analyses we had performed for the bilinguals and language learners suggested that there may be differences in performance. Looking at accuracy rates, bilinguals show a clear language switch cost, but no grammaticality effect (nor any effect of language); language learners show a grammaticality effect and a language effect, but no language switch cost. Proficiency overlapped among our participants, but it was controlled for, and we wished to directly test whether native-ness was affecting the data, especially in regards to grammaticality.

We also wished to see if proficiency itself was affecting grammaticality, rather than different the language profiles encompassed by participant type. We therefore fit a linear mixed-effects model including interactions with Spanish proficiency. We did not include an interaction with English proficiency as it was not significant in models conducted in Experiments 1 and 2, and many participants were at ceiling level. We likewise fit a linear mixed-effects model comparing grammatical predictability for same and mixed language trials including interactions with Spanish proficiency, to see if Spanish proficiency affected the amount of grammatical predictability found equally in same and mixed language trials.

Our final reason for looking at all the data combined was to see if code-switching use or sensitivity to code-switching constraints affected grammatical predictability. For these analyses, we did not include participant type as a factor; instead, we combined the data, and fit linear mixed-effects models using measurements of code-switching use and sensitivity to code-switching constraints as factors.

Question 1: Does Language 'Native-ness' Affect Grammatical Predictability?

The first variable we considered was whether or not language 'native-ness' played a role in participants showing grammaticality effects. In other words, we wished to see if language learners and bilinguals were performing differently in the experiment. If they were, native-ness is likely playing a role; if not, native-ness is not, and linguistic knowledge is the most likely factor affecting grammatical predictability.

In regards to accuracy, Experiments 1 and 2 had very different patterns. Language learners showed evidence that grammaticality and language but not language congruency affected accuracy rates, while in contrast, bilinguals showed evidence that language congruency but not language nor grammaticality affected accuracy rates. Predictably, the combined data showed main effects of grammaticality, language congruency, and language, and interactions with participant type, suggesting that native-ness is playing a role. Crucially, language learners were more sensitive to grammaticality than bilinguals. While we did not predict that bilinguals' accuracy would be insensitive to grammaticality, this finding underscores that language learners are indeed sensitive to grammaticality.

Focusing instead on reading time, we see that Experiments 1 and 2 had very similar patterns, with a grammatical predictability effect being found in bilinguals, and a strong trend to the same for language learners. To make sure that the patterns were more than superficially similar, we analyzed the data using participant type as a factor.

Unsurprisingly, we found the same pattern as in Experiment 1: participants read nouns following adverbs slower than those following determiners, and there was no interaction with either language nor with language congruency, showing language-independent grammatical predictability. Participants read nouns in mixed language word pairs slower than nouns in same language pairs, and this did not interact with language. Finally, participants were slower at reading words in Spanish than in English, which we had found previously in both language learners and bilinguals.

Participant type was not found to be significant, showing that bilinguals and language learners were performing similarly. Importantly, not only were the same patterns being found, but the magnitude of the main effects of interest were not significantly different; in regards to our main variables of interest, there was no difference in regards to grammaticality nor language switch cost. There was a strong trend for participant type to interact with language, as language learners showed a greater difference in their reading speeds between English and Spanish; however, this is foreseeable as language learners in general had a greater difference in the linguistic proficiencies of their two languages.

When comparing the grammatical predictability effect in same and mixed language trials separately, we found the effect size to be similar for both types of trial, further suggesting that participants show language-independent grammatical predictability. As we did not find less of grammatical predictability effect in mixed language pairs than same language pairs, this provides further evidence against the bilingual production model (de Bot, 1992), which would predict that grammatical predictability, if found in mixed language trials, would be less than that found in same language trials.

In other words, we found no difference between language learners and bilinguals, showing that the main effects of grammaticality, language congruency, and language were likely driven by all participants. Given these results, we can conclude no evidence of language 'native-ness' affecting grammatical predictability, as measured in the reading times.

Evidence that language learners show grammatical predictability. As we found no difference between language learners and bilinguals in the reading times, and language learners were more sensitive to grammaticality than bilinguals in the accuracy data, we can conclude that language learners are showing grammatical predictability, though we found only a trend when looking at the learner reading times directly. Furthermore, as we found no evidence of interactions with language or congruency, we can say that language learners are showing language-independent grammatical predictability, as do bilinguals. This provides further support for the shared syntax account (Hartsuiker & Pickering, 2008), which suggests that words are activated by any possible grammatical structure in which they could fit, regardless of language.

Language learners and the Revised Hierarchical Model. Having found evidence of grammatical predictability, we can discuss whether or not it provides support for the Revised Hierarchical Model (Kroll & Stewart, 1994). The most immediately relevant finding is that similar predictability effects were found regardless of the language condition of the word pair. As such, we did not find evidence of automatic translation to English occurring, as would be predicted by the RHM. If automatic translation were occurring, we would expect that Spanish-English word pairs would show similar predictability effects as English-English pairs, since the English word form would be activated by automatic translation at the first word. In contrast, we would expect that English-Spanish word pairs to be the least likely to show evidence of predictability, as there would be an additional processing cost due to the automatic translation occurring at the Spanish word. We did not find any differences in regards to predictability, and as such we cannot say we see evidence of automatic translation.

Though we cannot rule out the possibility of it occurring and it not affecting grammatical predictability, it seems unlikely. Indirect evidence suggests otherwise. Automatic translation is predicted to occur most often in less frequent words, such as adverbs, but not in frequent words, since the L2 learner would have knowledge of those words. In other words, we would expect that more automatic translation to be occurring in the first word of ungrammatical trials than in grammatical trials. In regards to automatic translation, the second word (noun) should not show a difference, as the second word had similar frequency regardless of the grammaticality of the word pair. Automatic translation would increase processing cost in that first word, which we are not measuring; however, it also effectively changes Spanish-English word pairs to English-English word pairs, which should lead to less language switch cost as compared to English-Spanish word pairs, which we did not find.

Though we did not find evidence for automatic translation, it does not mean that our study provides support against the RHM. The RHM has automatic translation occurring in the earlier stages of language learning, with learners forming a connection between the word form in their new language and their native language, but not yet a connection between the L2 word form and the concept it refers to. The automatic translation is therefore necessary so that the L2 learners can access the semantic meaning of the concept being referenced. Afterwards, as linguistic knowledge increases, direct connections are formed between the L2 word forms and the semantic meaning, and automatic translation does not occur as it is not necessary. Unfortunately, there is no direct way to test the exact connections a language learner possesses, leaving one to speculate as to whether a learner has direct connections or not.

Though we cannot state with certainty that our participants have formed direct connections between Spanish word forms and their semantic meaning, we can say they have all taken multiple Spanish language courses, some of which were taught in Spanish, showing that they had had time and experience to potentially form these direct connections. Lower proficiency speakers with less experience with Spanish would be less likely to have formed direct connections, and be more likely to show evidence of automatic translation.

Ultimately, our findings neither support nor refute the Revised Hierarchical Model. While it might be tempting to say that since we found no evidence supporting the model, our findings show evidence against the model, our sample was too limited to come to any conclusions. The RHM predicts that language learners' behavior should change as their proficiency increases, as words in the L2 no longer require the activation of the L1 counterparts for analysis. As our participants behaved similarly to bilinguals, it may be that our participants have already formed direct connections.

Question 2: Does Spanish Proficiency Affect Grammatical Predictability?

Participant type was not found to be significant in reading times, suggesting that bilinguals and language learners are behaving similarly in the experiment. However, it is unlikely that language learners have always shown grammatical predictability in Spanish, as at some point, they would not have the linguistic knowledge to do so. One possibility is that grammaticality predictability is moderated by Spanish proficiency, and no differences were found because there was overlap in Spanish proficiency between bilinguals and language learners (Figure 29).

Similar to previous models, we found main effects of grammatical predictability, language switch cost, and language. Spanish proficiency did not reach significance, but there was an interaction between Spanish proficiency and language, showing that Spanish proficiency affects the reading time of Spanish words. Importantly, we did not find evidence that Spanish proficiency moderated grammatical predictability, suggesting that all participants behaved similarly in regards to predictability. When looking at the model comparing the effect of Spanish proficiency on grammatical predictability in congruent and incongruent trials separately, we found no indication that Spanish proficiency affected grammatical predictability in either type of trial; instead, we found only evidence suggesting that the grammatical predictability effect was greatest in English-Spanish trials regardless of the proficiency of the participant.

One possibility is that all our participants had enough Spanish linguistic knowledge to form grammatical predictions in all language conditions. All the Spanish-English bilinguals had been speaking Spanish since early childhood (at least 15 years per participant), and the language learners had been studying Spanish for at least 5 years. Any differences in predictability due to proficiency may be too subtle to be found using the current methodology. To further look into the effects of proficiency, it would be advantageous to look at language learners who are less proficient in Spanish.

Question 3: Does Code-switching Use Affect Grammatical Predictability?

The next question considered was how code-switching use affected grammatical predictability. A possibility is that bilinguals show grammaticality effects due to previous experience with code-switching.

Language is acquired through experience; if code-switching functions as its own language, it is logical to assume that code-switching experience is how one learns the underlying grammatical rules. Additionally, more experience would lead to more effective parsing strategies, and grammatical predictability effects may be shown.

Similar to previous models, we found main effects of grammatical predictability, language switch cost, and language. Mean time spent code-switching with top conversational partners was not found to be significant, nor did it interact with grammatical predictability, language switch cost, or language. This pattern of results suggests that for our participants, mean time spent code-switching does not affect grammatical predictability. Such a finding does not correspond with a view that Spanish-English code-switching linguistic constraints are dependent on experience with code-switching.

Though code-switching experience appears to have no effect, these results are only suggestive; other factors may be affecting the results. First, we asked people to rate how much time they spent code-switching with their top conversational partners, which is not an objective measure; depending on a person's circumstances, this may define different amounts of time spent talking. For example, a person who speaks to many different people during the day may spend very little time code-switching, no matter how much time they spent code-switching with their top conversational partners.

Though we had no specific measure to control for this variability, we did ask people to estimate how much time they had spent in the last week code-switching. Bilinguals were found to code-switch more with their top conversational partners than they had the past week, though language learners were not. The difference among bilinguals and language learners may be reflective of a person's linguistic environment; most of our bilinguals were Heritage speakers at college who might code-switch more often when they are in an environment that includes more bilinguals. If recent experience plays a role, then it might add noise to the data.

Question 4: Does Sensitivity to Spanish-English Code-switching Grammar Constraints Affect Grammatical Predictability?

Finally, we looked at whether sensitivity to code-switching constraints affect grammatical predictability. Grammatical predictability is thought to occur because of the knowledge of the language; sensitivity to grammatical constraints is a way of measuring a person's linguistic knowledge. As we are interested in whether grammatical predictability persists across language switches, we looked at sensitivity to Spanish-English code-switching constraints.

To do so, we used participant's ratings of ungrammatical code-switching sentences as our measure of sensitivity. People who are not sensitive to code-switching constraints should find the ungrammatical sentences as acceptable as the grammatical sentences, while those who are should find the grammatical sentences more acceptable than the ungrammatical sentences. Otherwise stated, grammatical sentences should be found acceptable regardless of a person's sensitivity to linguistic constraints; ungrammatical sentences should only be found unacceptable by people who are sensitive to code-switching constraints. As we measured grammaticality using a 7-point Likert scale, we then have a range of grammatical sensitivity among our participants, with people who scored ungrammatical code-switching sentences lowest as most sensitive as sensitivity decreasing as acceptability rating increases. One indication that ungrammatical code-switching sentence rating is a valid measure is that bilinguals tended to rate ungrammatical code-switching sentences as lower than language learners; logically, bilinguals can be thought of as being most likely to be sensitive to code-switching constraints.

As in previous models, we found a main effect of grammaticality, with participants readings nouns faster following determiners than adverbs. This did not interact with language congruency, showing that the effect was found regardless of the language of words making up the word pair; in other words, this model also showed that grammatical predictability is language-independent.

We did not find a main effect of grammatical sensitivity (written as UCS [ungrammatical code-switch] in our tables and graphs); we did however, find an interaction between language congruency and grammatical sensitivity, with the language switch cost decreasing as sensitivity decreased. Interestingly, this was not due to the reading times of mixed language word pairs, but instead the reading times of same language pairs, with participants least sensitive to code-switching constraints being slower to read nouns in same language pairs than more sensitive participants. The interaction is not caused by language, with participants showing the same pattern regardless of the language of the second word. As most participants who rated the ungrammatical code-switching sentences were bilingual, this interaction is most suggestive of bilinguals and language learners showing differences in behavior. However, this is speculative, and more targeted research would have to be done in language switch cost in comprehension before any conclusions could be drawn.

Chapter 5

Experiment 3: Accounting for Grammatical Class Ambiguity

Introduction

For both monolinguals and bilinguals, highly predictable words are more quickly processed than uncommon or ungrammatical words during language comprehension. In our earlier experiments, we looked at whether grammatical predictability was affected by code-switches in both bilinguals and language learners, focusing on whether grammatical representations were shared across languages, and were neutral as to what caused the predictability effects. To do so, we assumed an idealized view of grammaticality; word pairs were considered either grammatical or ungrammatical. Predictability effects were expected in cases where the word pair formed a grammatical constituent, but not in cases where the word pair did not.

By doing so, we overlooked an important property of words. Specifically, that many words are ambiguous, which leads to ambiguity in regards to grammaticality. By word ambiguity, we are not referring only to the fact that words can have more than one meaning. Instead, we are focusing on the fact that these multiple meanings can be of more than one grammatical class. For example, the word 'dust' in English most often refers to fine particles of matter, a noun, but it also refers to the action of removing fine particles of matter from an area, as in cleaning (verb). Most common nouns in English can be read as verbs.

In previous experiments, grammaticality was defined in a binary fashion; a word pair was grammatical or not. However, recognizing that the same word form can refer to more than one grammatical class complicates the issue. Our grammatical word pairs were 'det-N', and our ungrammatical word pairs were 'adv-N'. An ambiguous word leads to the word pairs being able to be interpreted as 'det-V' and 'adv-V', which are ungrammatical and grammatical respectively, the opposite of what we were are claiming in regards to grammaticality!

In previous studies, we did take some measures to try to minimize the issue; for example, we tried to use words that were as unambiguous as possible. Not all words can be more than one grammatical class, and we tried to choose such words as possible. However, we were limited to choosing words that appeared early on in learning Spanish as a second language, which meant that our stimuli weren't completely unambiguous. Furthermore, this means limiting our stimuli to a small subset of language, which is not representative of natural English or Spanish.

In this experiment, to account for ambiguity, we did not treat grammaticality as a binary variable. Instead, we used word frequencies to calculate how often words appeared as a grammatical class, and then conditional probabilities to calculate how often that grammatical class either preceded or followed the other word of the word pair. By doing so, we were able to calculate more fine-grained, gradient measures. We calculated two such measures, separating the word pair into a lexical item and a category item. Specifically, we looked at the conditional probability of the first word given the grammatical category of the second word (w1ccProb) as the first measure, and the conditional probability of the second word given the grammatical category of the first word (w2ccProb) as our second measure.

The w1ccProb was calculated using the first word of the word pair and the grammatical class of the second word. It is the probability of the first word given the grammatical category of the second word. This measure accounts for sensitivity to lexical identity of the first word, and takes into account the ambiguity of the grammatical class of the second word. Ambiguity is not an issue in regards to the category of the first word since we are not making any claims as to its grammatical class in this calculation.

The w2ccProb is complementary to the w1ccProb. It was calculated using the second word of the word pair, and the probability of it being preceded by a word of the grammatical class of the first word. This measure accounts for the ambiguity of the grammatical class of the first word while not making any claims as to the grammatical class of the second word. This measure looks at how likely a word is given the previous grammatical category.

Though we do not make a claim that these measures substitute for grammaticality in general, they do allow for a probability-based account of words occurring together as specific word pairs. In the following section, we will discuss predictability in more detail, concentrating on current views of predictability.

Predictability Effects

The processing costs related to predictability have been explained by two broad classes of theories, *prediction-driven* (e.g., Altmann & Kamide, 1999; McRae et al., 2005; Pickering & Garrod, 2013; Rothermich & Kotz, 2013; Tanenhaus & Brown-Schmidt, 2008), and *integration-driven* (e.g., Frazier & Fodor, 1978; Jackendoff, 2002; Morris, 2006; Rayner et al., 2004). Prediction-driven theories assume that as we hear speech or read a sentence, we create probabilistic expectations of what word is to follow based upon prior words and our linguistic knowledge (e.g., Hale, 2001; Levy, 2008). An alternative and potentially complementary account is that processing costs arise because unpredictable words are difficult to integrate into our developing interpretations of a sentence. Less predictable words (including low frequency or ungrammatical words) may have a higher processing cost because they either take more effort to process (West & Stanovich, 1986). Separating processing costs due to prediction versus integration is difficult because the two theories often make the same empirical predictions (Kutas et al., 2011; Kuperberg & Jaeger, 2016). In the following sections, we will further discuss prediction-driven theories and integration-driven theories of predictability.

Integration-driven theories of predictability. Integration-driven theories of predictability can be thought of as showing a bottom-up view of processing. It is not that the sentence structure leads a reader/listener to differentially *expec*t a specific word/grammatical structure; rather that words conforming to the developing interpretation do not require extra processing costs. To put it in the context of our study, it is not that a participant expects a noun following a determiner; rather that a noun is easiest to process in that structure. Following an adverb, a noun is not easiest to process, which leads to increased processing cost—the differences in processing are not due to whether a noun is expected, only how difficult it was for the participant to interpret the word.

Prediction-driven theories of predictability. Prediction-driven theories of predictability are based upon the idea that during sentence interpretation, a person uses previously encountered information to expect some forms more than others. In contrast to the ideas posited by integration-driven theories, prediction is a top-down approach. A parser uses previous information to activate structures before they occur. Many prediction-driven theories are probability-based, using linguistic knowledge to anticipate specific words or grammatical structures (Levy, 2008). Whether or not the word ultimately fulfills the prediction or not is what leads to differences in processing costs. In other words, prediction-driven theories rely on how earlier words lead to pre-activate information, and since it takes less to activate a word/structure which is previously activated, it requires less processing cost. The greater ease in processing is what leads researchers finding predictability effects.

Pre-activation can be seen in a variety of contexts, often facilitating language processing. In cases of pre-activation of a semantic meaning, adding context can prevent the garden path effect in ambiguous sentences (Crain & Steedman, 1985). Semantic pre-activation effects can be very fine-tuned; highly constraining contexts can lead participants to expect specific words. After reading '*He caught the pass and scored another touchdown. There was nothing he enjoyed more than a good game of...'*, subjects expected the word 'football' and showed an N400 effect for words which were not 'football'. However, they showed a smaller N400 effect for within-category exemplars (e.g., ‘baseball’) as compared to between-category options (e.g., ‘monopoly’), though both are equally unlikely; shared conceptual features between ‘baseball’ and ‘football’ had been activated, moderating the size of the N400 effect (Federmeier & Kutas, 1999; Federmeier et al., 2002). Highly-constraining contexts can also lead to difficulties in processing. For example, readers report finding the sentence ‘*the astronomer married the star*’ as temporarily ambiguous; the word astronomer pre-activates the meaning of celestial-body for star though it does not contextually fit the word ‘marry’, leading to processing difficulties (Charniak, 1983).

Highly constraining contexts also lead speakers of grammatical languages to anticipate upcoming syntactic features. In Both Spanish (Wicha et al., 2004) and Dutch (van Berkum et al., 2005) speakers show an increased P600 when an article not corresponding to the expected word occurs (though it is grammatical, and another word could reasonably follow). Previous studies have shown that grammar violations are linked to increased P600 effects (Molinaro et al., 2011), but there was no grammar violations present in these studies; the disagreement is only in regards to the word the participant was expecting.

Though highly constraining contexts have been specifically linked to expectations affecting processing time, less constraining contexts do not seem to lead to very specific predictions. For example, Federmeier & Kutas (1999) found that reading a less constraining context such as *‘The patient was in critical condition and the ambulance wouldn’t be fast enough. They decided they would have to use the...’* leads a participant to expect the word ‘helicopter’, but it does not have such a high cloze probability as ‘football’ does in the example mentioned previously. When presented with an unexpected exemplar, participants do show a N400 effect. However, it is not modulated by whether the unexpected exemplar is within-category (‘plane’) or between-category (‘ferry’), suggesting that shared within-category perceptual features have not been activated by the context.

This is not to say that less constraining contexts have not been associated with predictability effects; instead, they have been unable to be distinguished as being due specifically to prediction. Longer reading times due to an unexpected syntactic category can be thought of as being due to predictability, but it is not possible to say whether it is because the incoming context led the parser to expect a specific grammatical category or whether it was easier to integrate a word that does not violate grammatical rules, i.e., the prediction vs. integration debate in predictability.

Integration vs. prediction debate. As we have seen, predictability effects can theoretically be caused by either prediction (word/grammatical structure being expected based on previous information) or by integration (differences in processing cost due to processing difficulty). It is difficult to distinguish the two because the predictions made are largely similar; expected words are ones that fall into a developing grammatical structure, which is the easiest to process. It is important to note that one concept does not exclude the other; potentially, both prediction and integration may be factors in predictability effects.

Current Experiment

The current experiment seeks to focus on several aspects that our previous studies did not account for. First, we are using a wider range of words, not limiting words to those a student learns during their first year of learning Spanish; thus Experiment 3 contains a more natural distribution of words that occur in language. More importantly, we are accounting for grammatical class ambiguity, instead of treating grammaticality as a binary variable. To do so, we are using two probability-based measures to capture grammaticality. Doing so allows us to account for the fact that grammatical class ambiguity leads to overall grammatical ambiguity, i.e., a word pair may be read as both grammatical and ungrammatical depending on how a lexical item is interpreted.

Hypotheses. As in the previous experiments, we expect to find known word effects, such as frequency, length, and spillover effects. We also expect a practice effect, as participants speed up as the experiment goes on. Since this experiment is broken up into two sessions instead of one, it is likely that a session effect will be seen (participants will be faster the second session, since they are already familiar with the task).

A language switch cost is expected, with words following a language switch read slower than words not following a language switch. A language effect is likely, as we are recruiting in Michigan, and most participants are educated primarily in English, and therefore more practiced at reading English (they are likely more English-dominant). In other words, words in English are likely to be read faster than words in Spanish.

Of most interest, we wish to see whether the conditional probability measures are associated with a predictability effect. Specifically, we are most interested in the w2ccProb, the probability of the second word given the grammatical class of the first word. Our null hypothesis is to not see any effect, suggesting that no predictability effect is found; our alternate hypothesis is to find a predictability effect. In other words, as the w2ccProb increases, we expect to see faster reading speeds. We make no specific predictions in regards to the w2ccProb, as it is included mainly to account for sensitivity to the lexical identity of the first word.

Method

Participants

Forty fluent Spanish-English bilinguals from the Ann Arbor area completed the experiment. Participants received $14 base pay as well as bonuses based on speed and accuracy, for a total average payment of $26.13 (range: 21.99 : 29.36, sd = 1.57).

The mean age of the participants was 24 years (range: 18 : 36). Most participants learned both languages from an early age (mean AOA of Spanish 0.7 years, mean AOA of English 5.5 years) and rated themselves as speaking mostly English (76% of the time) and reading mostly English (82% of the time) in the 2 weeks prior to the study. Participants rated their speaking skills as 6.6 (Spanish) and 6.5 (English), and their reading skills as 6.1 (Spanish) and 6.4 (English), on a 7 point scale.

Materials

Word stimuli. As in the two previous experiments, Experiment 3 used word pairs that were either grammatical, determiner-noun (det-N), or ungrammatical, adverb-noun (adv-N). Different words were used than in the previous studies, to make sure that grammaticality would occur over a different set of stimuli.

English words and frequencies (per million) were drawn from the Corpus of Contemporary American English (Davies, 2008), and Spanish words and frequencies from the Corpus del Español (Davies, 2002). Each word was matched with its translation. Because Spanish determiners have a gender, English determiners were more frequent (average log frequency = 6.6) than Spanish determiners (average log frequency = 4.3). English adverbs were slightly more frequent (average log frequency = 2.2) than Spanish adverbs (average log frequency = 1.7). Nouns had similar frequencies in both languages; English (average log frequency = 3.8), Spanish (average log frequency = 3.5).

Nonwords were created to follow the phonotactic rules of the respective languages. In total, there were 200 English nonwords and 200 Spanish nonwords. Each participant saw each nonword once during the experiment.

Word lists were divided so as to have equal numbers of language-language combinations, equally spread through each type of critical word pair. For example, each participant saw 60 det-N trials, 15 in each language combination. For simplicity, we will refer to total numbers of word pair types shown rather than specifying the number per language combination condition.

There were a total of 60 det-N and 60 adv-N critical trials. To prevent participants from anticipating that determiners would always lead to a grammatical construction, there were also 60 fillers that were det-prep, det-det, or det-conj, and 120 det-nonword pairs. To prevent participants from anticipating that adverbs would always lead to an ungrammatical construction, there were 60 adv-V fillers, and 120 adv-nonword pairs. Participants also saw an additional 160 filler pairs, the majority being N-det and adj-adj pairs, but also including 35 other types such as adv-adj, det-adj, det-V, adj-adv, adj-det, adj-V, N-adj, N-N, V-det, V-adj, V-V. Other than the nonword pair types already mentioned, participants also saw 120 nonword-N pairs, and 40 nonword-adj pairs.

Conditional probabilities. To account for the fact that the same word form may be ambiguous as to category, we did not treat word category as absolute for each word pair. For example, the word 'dog' may be interpreted as a noun, referring to the canine, or as a verb, referring to the act of following someone. If 'dog' were used as a noun in one of our critical pairs, it could be argued that it could be interpreted grammatically in both cases: as det-N, or adv-V. Similarly, it could be interpreted as ungrammatical in both cases: det-V, adv-N. We have to consider that grammaticality cannot be considered to be absolute for all word pairs; instead, it can be on a sliding scale. Instead of treating grammaticality as a binary variable, we instead we operationalized two types of probabilistic effects based on the word pair: w1ccProb (P(Word1|Category2)) and w2ccProb (P(Word2|Category1)).

In monolingual studies, the w2ccProb of a given word pair might be estimated as P(Word2|Word1). Similarly, the w1ccProb would be estimated as P(Word1|Word2). These two conditional probabilities need not be the same: consider a word pair like 'suede shoes', where the first word highly constrains the second, but the second word could have followed many other preambles. Thus, the w2ccProb for 'those fish' means the probability of the word 'those' being followed by the word 'fish'. In other words, the probability of the word 'fish' given the word 'those'. It would be measured by taking the probability that 'those' was immediately followed by 'fish' and dividing it by the probability of 'those' being followed by any word (Figure 52).

$$P('fish'|'those') = \frac{freq('those'+'fish')}{freq('those' + \ any\ word)} = \frac{0.24}{956.68} = 2.51 * 10^{-4}$$

Figure 52: Conditional probability with two lexical items example: Probability of 'fish' given 'those'. Frequencies (occurrences per million) were found using the CoCA.

To our knowledge, there does not exist a Spanish/English code-switching corpus large enough to estimate these quantities. Instead, we used grammatical category to compute the proportions, combining specific word frequencies and category frequencies. Rather than the conditional probability of a word occurring given the other word, we calculated the conditional probability of a word occurring given the other word's grammatical category. To return to our example of 'those fish', the w2ccProb would be changed to mean the probability of the word 'fish' given a determiner, and the formula would change accordingly. It would be measured by taking the probability that a determiner is followed by the word 'fish' and dividing it by the probability of a determiner being followed by any word (Figure 53).

$$= \left(\frac{freq(\mathrm{det} + 'fish')}{freq(\mathrm{det} + any\ word)}\right) = \left(\frac{31.93}{163242.63}\right) = 1.96 * 10^{-4}$$

Figure 53: Conditional probability of the second word of a word pair given the (unambiguous) grammatical category of the first word example. P('fish'|determiner) is the probability of the word 'fish' given a determiner. + means combination of two words in that order.

Similarly, the w1ccProb would be the probability of the word 'those' given a noun. It would be measured by taking the probability that 'those' was immediately followed by any noun and dividing it by the probability of a noun being preceded by any word (Figure 54).

$$P('those'|\text{noun}) = \frac{freq('those' + \text{noun})}{freq(any\ word + \text{noun})} = \frac{451.54}{304063.91} = 1.49 * 10^{-3}$$

Figure 54: Conditional probability of the first word of a word pair given the (unambiguous) grammatical category of the second word example. P('those'|noun) is the probability of the word 'those' given that the category of the second word is a noun. + means combination of two words in that order.

Mixed word-category frequencies were found using the CoCA for English and the Corpus del Español for Spanish. As stated before, no code-switching corpus exists extensive enough to calculate conditional probabilities for code-switches using specific words. Therefore, mixed word-category frequencies were found using the corpus corresponding to the language of the word. For example, if we were looking at 'el lake', then the Corpus del Español would be used to find frequencies such as 'el' + noun, but the CoCA would be used to find frequencies such as det + 'lake'.

The above formulas gives us the theoretical values for 'fish' following any determiner and for 'those' preceding any noun. For the purposes of this study, we are generalizing it to the word pair 'those fish'. However, in doing so, we are making the assumption that 'those' can only occur as a determiner and 'fish' as a noun. We must take into account that lexical items can be ambiguous in respect to grammatical category. We must take into account that lexical items are ambiguous as to category; if not, we ignoring the fact that 'fish' can occur as both a noun and a verb. This is especially important as we are interested in grammaticality; by syntactic rules, det+N is grammatical and adv+N is ungrammatical. However, if 'fish' can occur as both a noun and a verb, then grammaticality is not clear cut, as adv+V is grammatical as well.

Since words may occur as more than one part of speech, our formulas must take into account several factors. First, we must determine how many syntactic categories each word of a given word pair can be interpreted as, and the frequency that they do so in relation to their frequency in the language. This allows us to find the ratio that the word exists as that part of speech in the language. To go back to our example, we have the word pair 'those fish', and we wish to know the conditional probability of 'those' given a word of the syntactic category of Word 2, in this case, 'fish'. 'Fish' occurs as both a noun and a verb, so we must first calculate the percentage of times it does so for each syntactic category. Next, we calculate the conditional probabilities of *P*(*'those'*|noun) and *P*(*'those'*|verb), using the same formulas as we did in Figure 54. We then multiply the conditional probability to the percentage of times that 'fish' occurs as that syntactic category. So for *P*(*'those'*|noun), we multiply it by the percentage of times that 'fish' occurs as a noun overall. Finally, we add the multiplied values to find the probability of 'those' given a word that is of the syntactic category of the word 'fish'. Figure 55 shows the general formula for w1ccProb; Figure 56 shows the formula as it is applied to the example 'those fish'.

$$\sum_{i=part\ of\ speech\ for\ word2}^{n=all\ possible\ parts\ of\ speech\ word2\ can\ be} \frac{freq(word1 + \ word.[i])}{freq(any\ word\ +\ word.[i])} \times \frac{freq(word2.[i])}{freq(word2)}$$

Figure 55: Formula for conditional probability of the first word of a word pair given the grammatical category of the second word [w1ccProb]. '+' means combination of two words in that order, and using the notation 'x.[y]' means that word x must be part of speech y. For our example of 'those fish', word1 is 'those', word2 is 'fish', and *i*= ('noun', 'verb'). Word.[*i*] means any word that is of syntactic category *i*, and word1.[*i*] refers to the frequency of 'those' occurring as *i*.

$$= \left(\frac{freq('those' + \text{ noun})}{freq(any\,word\; + \text{ noun})} \quad \times \quad \frac{freq('fish'.\,[\text{noun}])}{freq('fish')}\right)$$

$$+ \left(\frac{freq('those' + \text{ verb})}{freq(any\,word\; + \text{ verb})} \quad \times \quad \frac{freq('fish'.\,[\text{verb}])}{freq('fish')}\right)$$

$$= \left(\frac{451.54}{304063.91} \quad \times \quad \frac{111.79}{119.30}\right) + \left(\frac{117.78}{207896.08} \quad \times \quad \frac{7.49}{119.30}\right) = 1.43 * 10^{-3}$$

Figure 56: Worked-out example showing w1ccProb formula applied to a word pair; second word is ambiguous in terms of category: 'those fish'. 'Fish' can occur as a noun and a verb, so it is the probability of 'those' being followed by a noun or a verb proportionally to the amount of times 'fish' occurs as each category.

To calculate the w2ccProb, we use similar techniques, only instead of the possible syntactic categories that Word 2 can be interpreted as, the possible ambiguity is due to the possible word categories that Word 1 can be interpreted as. To return to our example a final time, we would look at the possible syntactic categories that 'those' may occur as. 'Those' occurs as a determiner the overwhelming majority of the time, so in this case, no extra syntactic categories have to be taken into account for Word 1. Figure 57 shows the general formula for w2ccProb; Figure 58 shows the formula as it is applied to the example 'those fish'.

$$\sum_{i=part\ of speech\ for\ word1}^{n=all\ possible\ parts\ of\ speech\ word1\ can\ be} \frac{freq(word.[i] + word2)}{freq(word.[i]\ + any\ word)} \times \frac{freq(word1.[i])}{freq(word1)}$$

Figure 57: Formula for conditional probability of the second word of a word pair given the grammatical category of the first word [w2ccProb]. '+' means combination of two words in that order, and using the notation 'x.[y]' means that word x must be part of speech y. For our example of 'those fish', word1 is 'those', word2 is 'fish', and *i*= ('determiner'). Word.[*i*] means any word that is of syntactic category *i*, and word1.[*i*] refers to the frequency of 'both' occurring as *i*.

$$= \left(\frac{freq(\det + 'fish')}{freq(\det + any\ word)} \times \frac{freq('those'.[\det])}{freq('those')}\right)$$
$$= \left(\frac{31.93}{163242.63} \times \frac{1078.62}{1078.65}\right) = 1.96 * 10^{-4}$$

Figure 58: Worked-out example showing w2ccProb formula applied to a word pair; first word is unambiguous in terms of category: 'those fish'. 'Those' occurs overwhelmingly as a determiner, with other categories it can be occurring so few times (0.03 per million) that they do not affect the final value of the w2ccProb and so are not shown in the example.

It is important to note that using abstract grammatical categories like noun and determiner in our computations results in conditional probabilities akin to gradient, lexicalized grammatical knowledge about word pairs. Det-N pairs can be generalized as having greater w1ccProb and w2ccProb values on average than adv-N pairs; however, the conditional probabilities allows us to compare how specific words affect 'grammaticality'. For our specific word pairs, the conditional probabilities are correlated, $r(999) = 0.357$, $p < .001$, with det-N pairs having both higher w1ccProb and w2ccProb values. When comparing det-N and adv-N pairs, the difference is more pronounced for the w1ccProb than the w2ccProb, as there is a greater overlap in values for w2ccProb among the different pairs. Figure 59 shows the relationship between w1ccProb and w2ccProb for the word pairs used in the experiment.

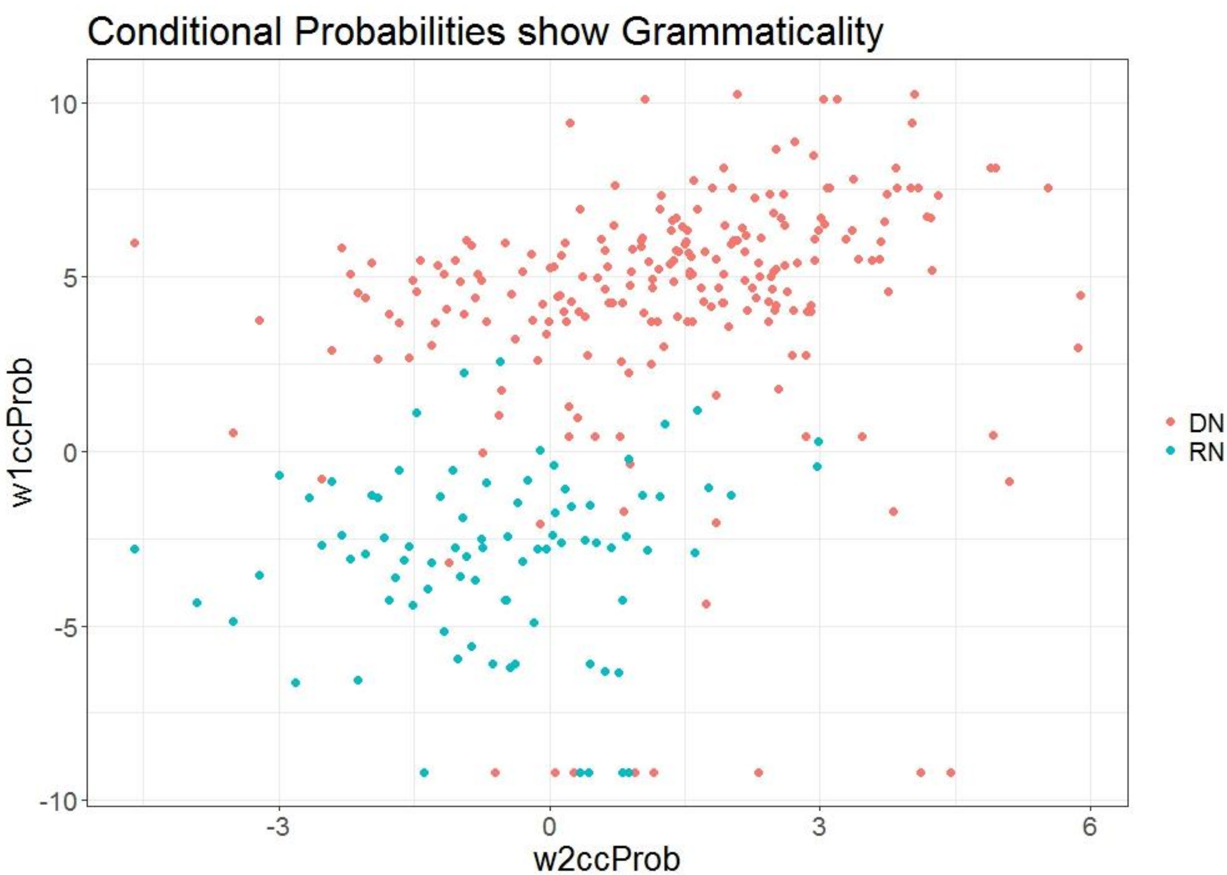


Figure 59: Correlation between w1ccProb and w2ccProb for word pairs used in Experiment 3. w1ccProb and w2ccProb are correlated for both det-N and adv-N word pairs; det-N pairs have both higher w1ccProb and w1ccProb values than adv-N pairs. Low w1ccProb values in det-N pairs are due to determiners with very low probabilities (might have appeared once in lexicon) and are not/rarely followed by words the category of the second word.

Procedure

Participants took part in two hour-long sessions, usually spaced seven days apart (range: 5 days : 25 days, median 7 days). Participants completed a list lexical decision task while being eye-tracked using an Eyelink 1000 eye-tracker set 2 feet from the screen each session. Before being eye-tracked, participants were calibrated using the Eyelink calibration software.

Each trial started with a single point drift correction in the center of the screen. Participants then focused on a fixation cross on the left side of the screen which automatically triggered the appearance of a two-word list followed by another fixation cross, all in 16-Pt Courier font. Participants were instructed to view both words, fixate on the rightmost cross, and indicate their response using the keyboard. Participants pressed a button indicating 'yes' if *both words* existed in either English or Spanish, and a different button otherwise. Instructions were presented in both English and Spanish. Responses were only accepted while fixated on the rightmost cross. The crosses were included to promote natural left to right reading patterns and to make it more likely that gaze duration on each word was due to reading, not making the lexical decision.

$$Bonus\ for\ correct\ answers = \frac{4250 - reaction\ time\ (ms)}{150}$$

$$Penalty\ for\ incorrect\ answers = 50 - \frac{4250 - reaction\ time\ (ms)}{150}$$

Figure 60: Formula for calculating reward/penalty in Experiment 3. If a participant takes 1250ms and answers correctly, the bonus would be 20pts. If the participant answers wrong at 1250ms, the penalty would be -30pt. If reaction time is greater than 4250ms, the bonus is 0pt, and penalty is 50pts.

Participants received points for correct answers and lost points for incorrect answers, following the formula given in Figure 60, prompting participants to follow a consistent speed-accuracy tradeoff. Feedback was given after each trial, stating correctness, time taken to answer, and points gained or lost. Every twenty trials participants were shown how many total points they had accumulated during the course of the experiment. Points were converted to dollars by dividing by 1000; so if a participant finished with a total of 6000 points, they would receive a bonus of $6.00.

Participants completed a total of 800 test trials, 400 per session, preceded by 8 practice trials each session. Stimuli were not repeated either within or across sessions. Half of the trials were 'yes' trials. Each participant was shown one of eight stimulus lists for a total of five participants per list.

Scoring

For our analyses, we looked only det-N and adv-N trials. In other words, that means we looked at word pairs in which Word 1 could be analyzed as a determiner followed by Word 2 being analyzed as a noun, and ones in which Word 1 could be analyzed as adverb followed by Word 2 being analyzed as a noun. Though the experiment was designed to contain 60 det-N and 60 adv-N pairs, this led to some trials which had originally been designated as filler trials being analyzed in the final dataset. In total, each participant was exposed to an average of 64 trials which could be classified as det-N and 62 which could be classified as adv-N. (Exact number varied by as much as 4 trials more or less of a type per list; each participant was shown one of a possible 8 lists.) For convenience, we will refer to det-N trials as grammatical and adv-N trials as ungrammatical; however, for the purpose of analysis, grammaticality was characterized by the w1ccProb and w1ccProb conditional probability measures for that particular word trial.

Other types of trials and 'No' trials were not analyzed. Due to the fact that mean reading times varied by participant (range: 196ms : 559ms), outliers were calculated by participant. Words with reading times less than 50 ms or more than three standard deviations from that participant's mean reading time were dropped from analysis (1.91% of trials), leaving 4765 total trials used in the analyses.

We performed several sets of analyses. We first conducted an accuracy analysis. For the accuracy analysis, we calculated each participant's average percentage correct (APC) by language X congruency x grammaticality, for a total of eight values per participant. Since the conditional probability measures are continuous, we used word pair type (det-N and adv-N), to sort the trials based on grammaticality (as in Experiments 1 and 2). Each percentage was calculated by dividing the number of correct trials by the total number of correct and incorrect trials in the particular language X congruency x grammaticality condition (trials that had been deemed outliers were not included in the calculation). In total, there were 320 values taken into the model, 8 per participant.

We also analyzed the first pass reading time of the second word as our primary measure. Only correct 'yes' grammatical and ungrammatical trials were analyzed, lowering the total number of trials analyzed to 4402. We also analyzed the probability of whether or not a participant made an eye movement regression from the second word back to the first word; as in the reading time analysis, only correct trials were scored. Trials that included an eye movement regression were scored as a 'hit', with a value of '1', and trials without an eye movement regression were scored as a 'miss', with a value of '0'. A final analysis looked at the first pass reading time of correct trials without regressions, lowering the total number of trials to 4074.

Results

Lexical Decision Task

Accuracy analysis. Average accuracy was 92.38% for critical trials, with highest accuracy for English-English det-N trials (98.11%), and lowest for Spanish-Spanish det-N trials (88.55%). Table 25 shows the accuracy rates for the different types of trials. We were unable to fit a full parsimonious logistic regression model that accurately represented the data, so we instead fit a linear mixed-effects model to the APC using the lme4 package in R (Bates et al., 2015; R Core Team, 2015). Since we are looking at average accuracies across trials, we were unable to include trial specific conditional probabilities per word pair to show grammaticality. Instead, the average w1ccProb and average w1ccProb were calculated for the trials that were used to make up each condition per participant. Average conditional probability values formed a distribution similar to that of the singular conditional probability values (Figure 59), capturing the same relationship (Figure 61).

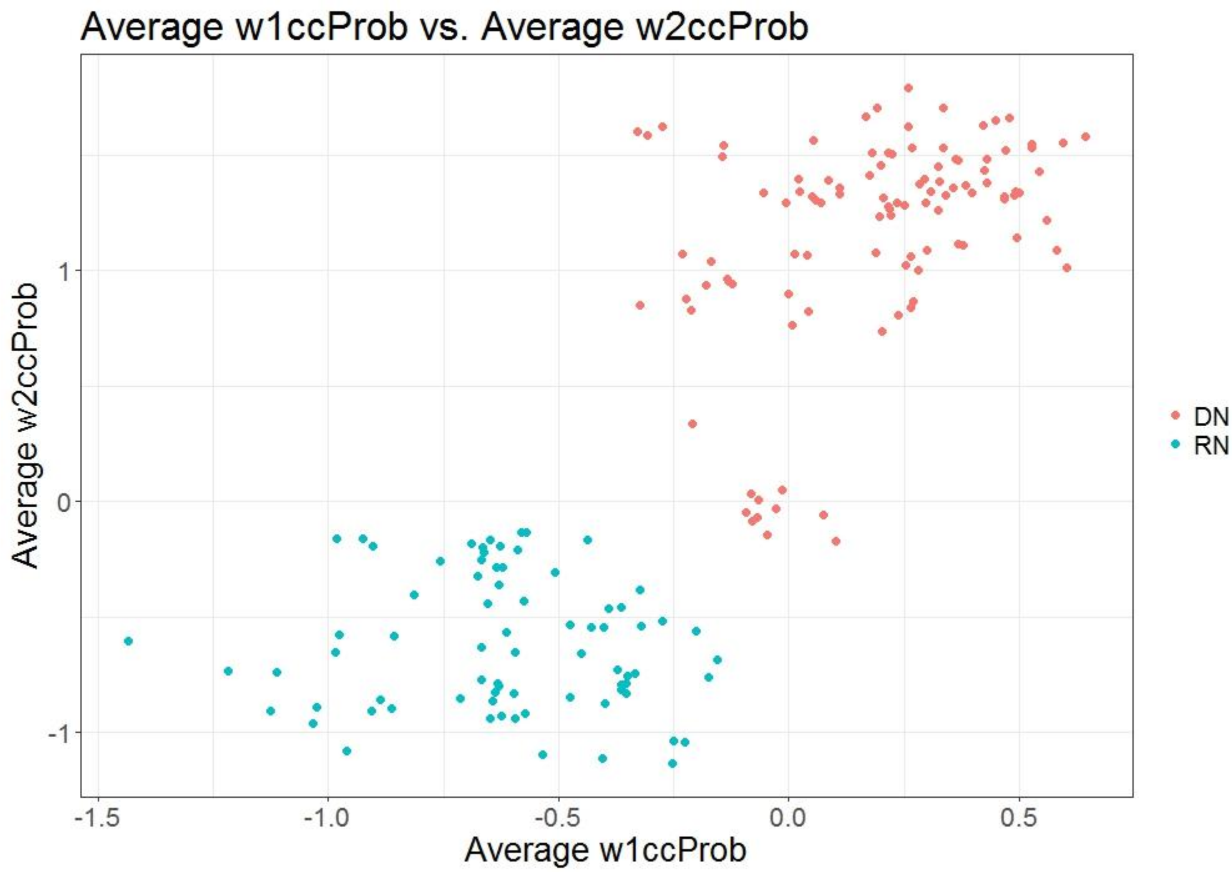


Figure 61: Correlation between average w1ccProb and average w2ccProb show distinct groupings for det-N and adv-N word pairs; captures the same relationship as singular values.

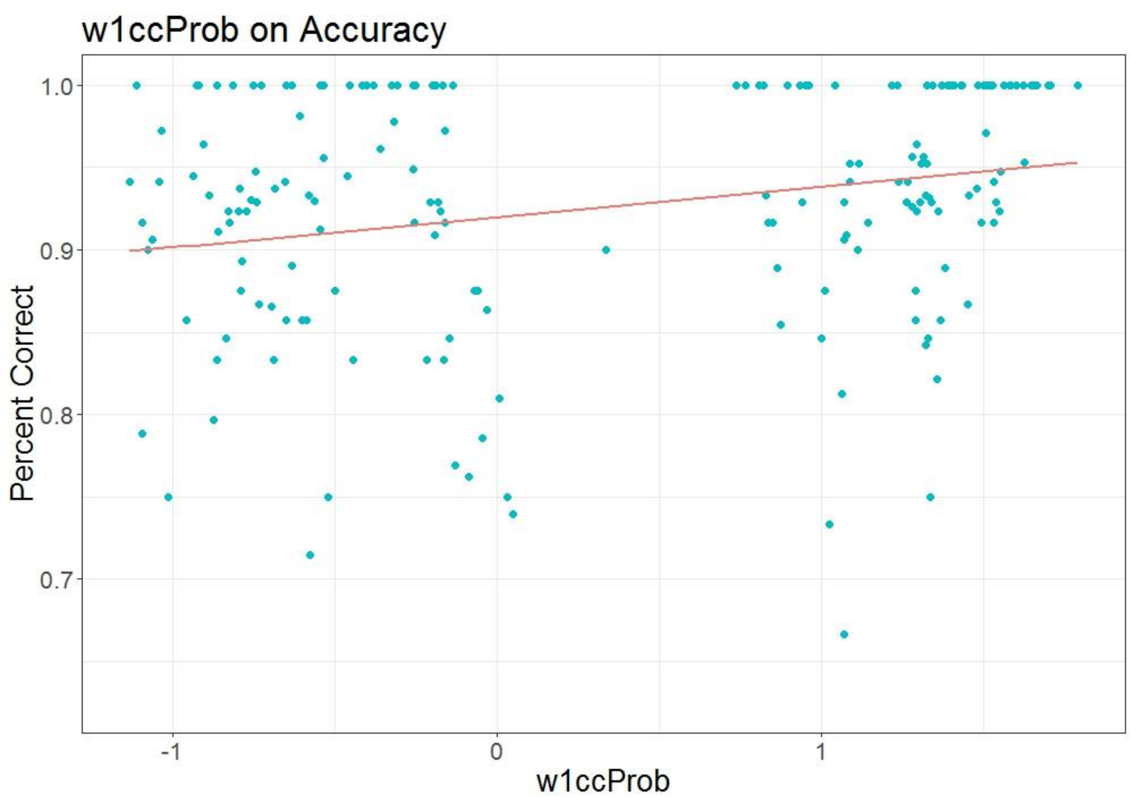
The model included fixed variables for average w2ccProb, average w1ccProb, language of the second word, whether or not the trial was a same or mixed language trial (congruency), all possible interactions between the variables, and a random variable for subject. Variables were centered before being added to the model.

As in previous studies, we followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subject, an intercept, and slope for language of the second word (Figure 62).

```
Model <- glmer(APC ~
             WordLanguage*LanguageCongruence*AvgW1ccProb*AvgW2ccProb +
              (1 |subject) +
                (0 + WordLanguage |subject)
             data=bilingual, REML = F)
```

Figure 62: Linear mixed-effects model for accuracy analysis in Experiment 3.

A main effect of w1ccProb was found, $\beta = 0.036$, $t(302) = 3.985$, $p < .001$, showing that participants increased in accuracy as w1ccProb increased. Similarly, a main effect of w2ccProb was found, $\beta = -0.053$, $t(302) = -2.471$, $p = 0.01$, with participants increasing in accuracy as w2ccProb increased. Both these effects show that accuracy improves as grammaticality improves (Figure 63).



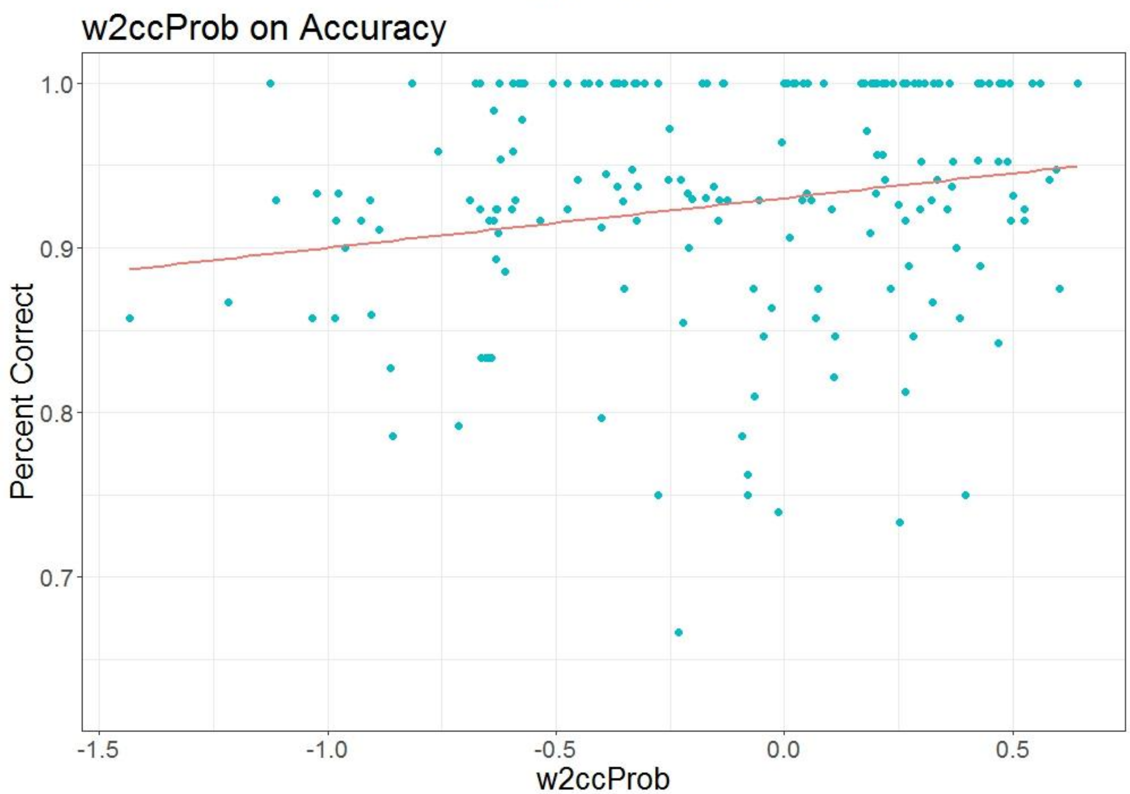


Figure 63: Relationship between conditional probabilities and accuracy. The top graph shows average w1ccProb on accuracy, and the bottom graph shows average w2ccProb on accuracy. As the conditional probability increases, so does accuracy.

Three-way interactions were also found between average w1ccProb, language, and congruency (β = -0.100, $t(302)$ = -2.645, p = 0.04), as well as between average w2ccProb, language, and congruency (β = 0.176, $t(302)$ = 2.046, p = 0.008). There is also a trend for a four-way interaction between language of the second word, congruency, average w1ccProb, and average w2ccProb, β = -0.129, $t(302)$ = -1.706, p = 0.09.

The fact that similar effects are found for both average w1ccProb and average w2ccProb are not surprising, as these values are highly correlated, $r(318)$ = 0.82, p <.001. The three-way and four-way interactions found are likely driven by the strong effect between language of the second word and language congruency (Figure 64). Differences in the rate at which accuracy improves as grammaticality improves in the different conditions leads to interactions involving the conditional probabilities.

Trials with the second word in English were answered more accurately than trials with the second word in Spanish, β = -0.047, $t(302)$ = -2.120, p = 0.03. There was no main effect of language congruency, but there was an interaction between congruency and the language of the second word, β = 0.154, $t(302)$ = 3.560, p < .001. This is because our participants were most likely stronger in English than in Spanish, disguising the effect of a language switch cost. Words in English were more easily identified as being real words as compared to ones in Spanish. Participants were more accurate in English-English trials as compared to Spanish-English trials due to being stronger in English and a language switch cost. English-Spanish trials and Spanish-Spanish trials are equally accurate, due the switch cost being mitigated by greater accuracy in English (Figure 64).

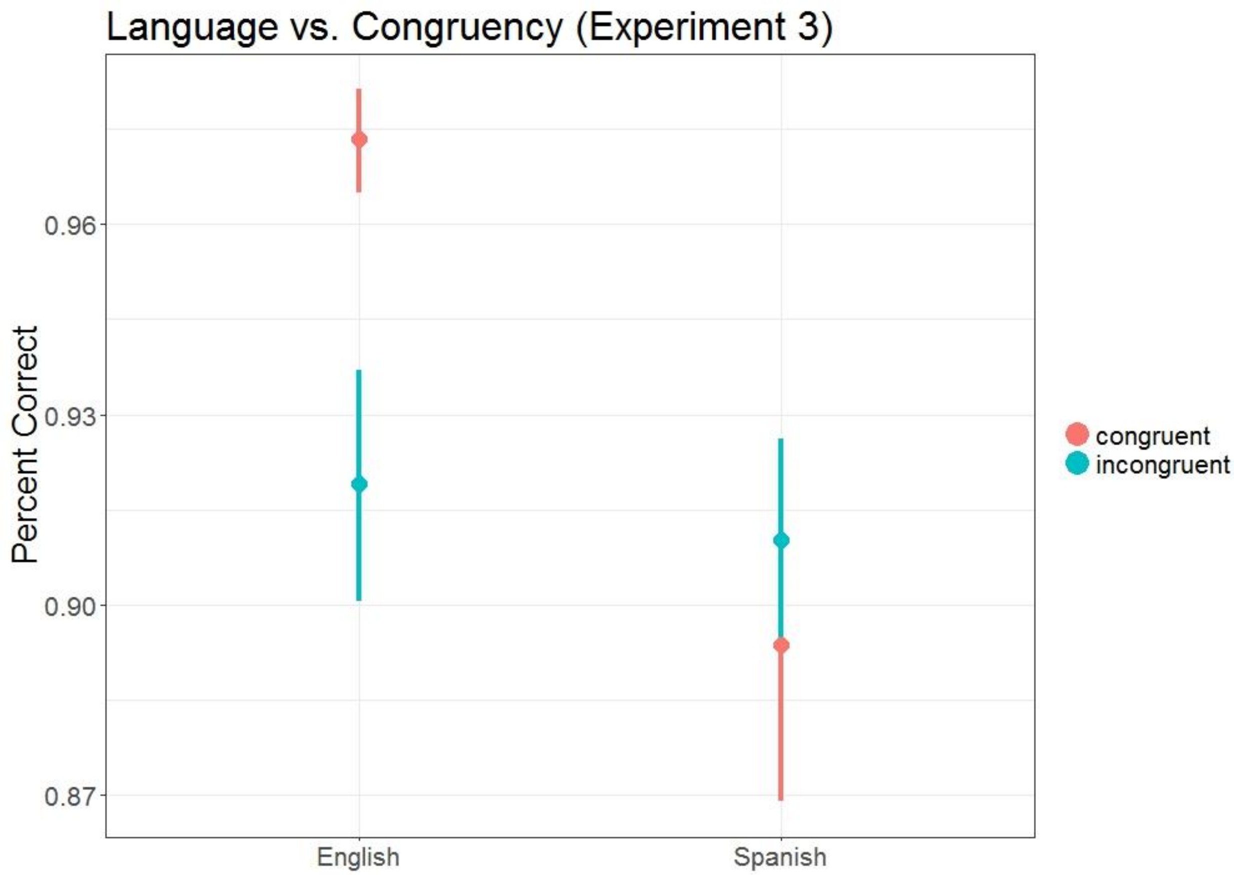


Figure 64: Congruency x language interaction for accuracy analysis in Experiment 3. Language refers to the language of the second word; error bars represent the standard error of the mean.

Table 25 shows accuracy rates by grammaticality (as defined as det-N and adv-N), language condition, and congruency. For each row, the topmost number is the overall percentage correct per condition (not separated by grammaticality), followed by the percentage correct as divided by grammaticality (det-N is grammatical, adv-N is ungrammatical). Abstract grammatical category is highly correlated with each conditional probability measure (with average w1ccProb, $r(318) = 0.91$, $p < .001$; with average w2ccProb, $r(318) = 0.84$, $p < .001$). However, abstract grammatical category fails to capture as much variance as the conditional probabilities, suggesting that using the more fine-grained measure leads to a more accurate representation of the data. When comparing a similar linear model that replaces the conditional probability values with a single binary grammaticality variable to the model shown in Figure

62, the model with a binary grammaticality variable is significantly different and has more deviance than the model with the conditional probability variables ($\chi^2(8, N = 320) = 27.68, p < .001$).

Table 25: Average accuracy rates in Experiment 3.

	Percent Correct	
English-English	97.07%	
	det-N	98.11%
	adv-N	96.23%
Spanish-English	91.89%	
	det-N	92.25%
	adv-N	91.59%
English-Spanish	91.31%	
	det-N	93.07%
	adv-N	89.06%
Spanish-Spanish	89.27%	
	det-N	88.55%
	adv-N	90.17%
Same Language	93.17%	
	det-N	92.80%
	adv-N	93.54%
Mixed Language	91.59%	
	det-N	92.71%
	adv-N	90.44%
Word 2 (English)	94.50%	
	det-N	95.18%
	adv-N	93.95%
Word 2 (Spanish)	90.30%	
	det-N	90.83%
	adv-N	89.61%
ALL	92.38%	
	det-N	92.75%
	adv-N	92.01%

Reading Time Analysis. We fit a linear mixed-effects model to the first pass reading time of the second word of correctly answered 'yes' det-N and adv-N trials using the lme4 package in R (Bates et al., 2015; R Core Team, 2015). To be included in the analysis, the second word must have been fixated upon; across the critical trials, no trials were discarded due to the participant not fixating on the second word.

Trials without a fixation on the first word were included, which accounted for 104 trials (2.36%). Skipping the first word did not vary by language condition or congruency, but det-N trials were more likely to be skipped than adv-N trials, with a total of 12 adv-N trials skipped as compared to 92 det-N. The fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, session number, and our main variables of interest: language of the second word, language congruency, w1ccProb, and w2ccProb, as well as all interactions between those four variables. All variables were centered before being added to the model.

Random effects were taken into account for both item and subject. As in Experiments 1 and 2, subject refers to the participant, and item refers to the second word; no participant was exposed to the same item more than once. As before, we followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for word length of the second word, word frequency of the second word, language of the second word, language congruency, session number, and trial number; for items, a random intercept, and slopes for w1ccProb, language congruency, and trial number (Figure 65).

```
Model <- lmer( FPRT ~ TrialNumber  +
            SessionNumber +
            WordLength +
            LogWordFrequency +
            SpilloverLength +
            LogSpilloverFrequency +
            LanguageCongruence*WordLanguage*w1ccProb*w2ccProb +
            (1 |subject)
              (0 + SessionNumber |subject) +
              (0 + WordLength |subject) +
              (0 + LogWordFrequency |subject) +
              (0 + WordLanguage |subject) +
              (0 + LanguageCongruence |subject) +
              (0 + TrialNumber |subject) +
            (1 |item) +
              (0 + w1ccProb |item) +
              (0 + LanguageCongruence |item) +
              (0 + TrialNumber |item),
            data=bilingual, REML=F)
```

Figure 65: Linear mixed-effects model for reading time analysis in Experiment 3.

Of our main variables of interest, w2ccProb was found to be significant, β = -9.429, $t(4369)$ = -2.436, p = 0.02, with nouns being read faster as the conditional probability that they were preceded by a word of the category of the first word increased (Figure 66).

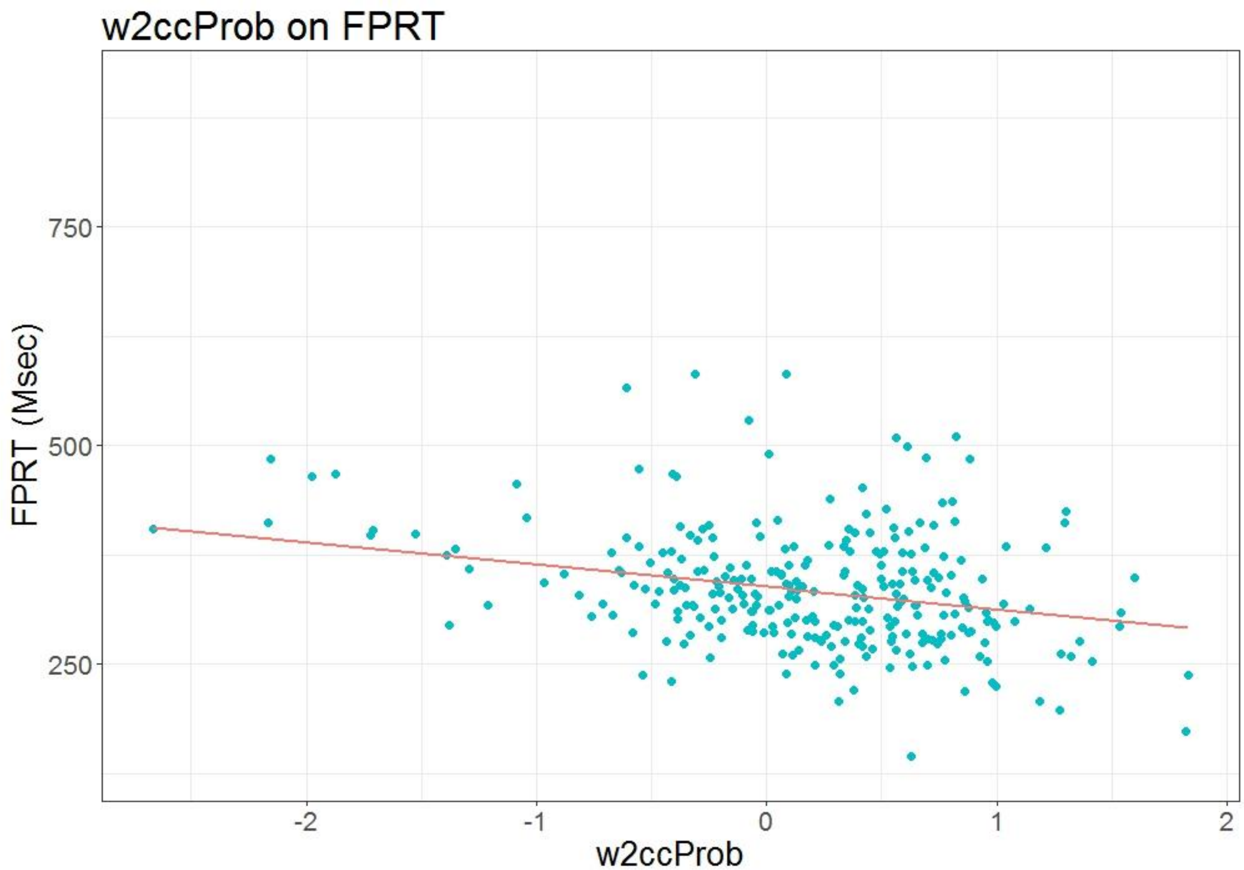
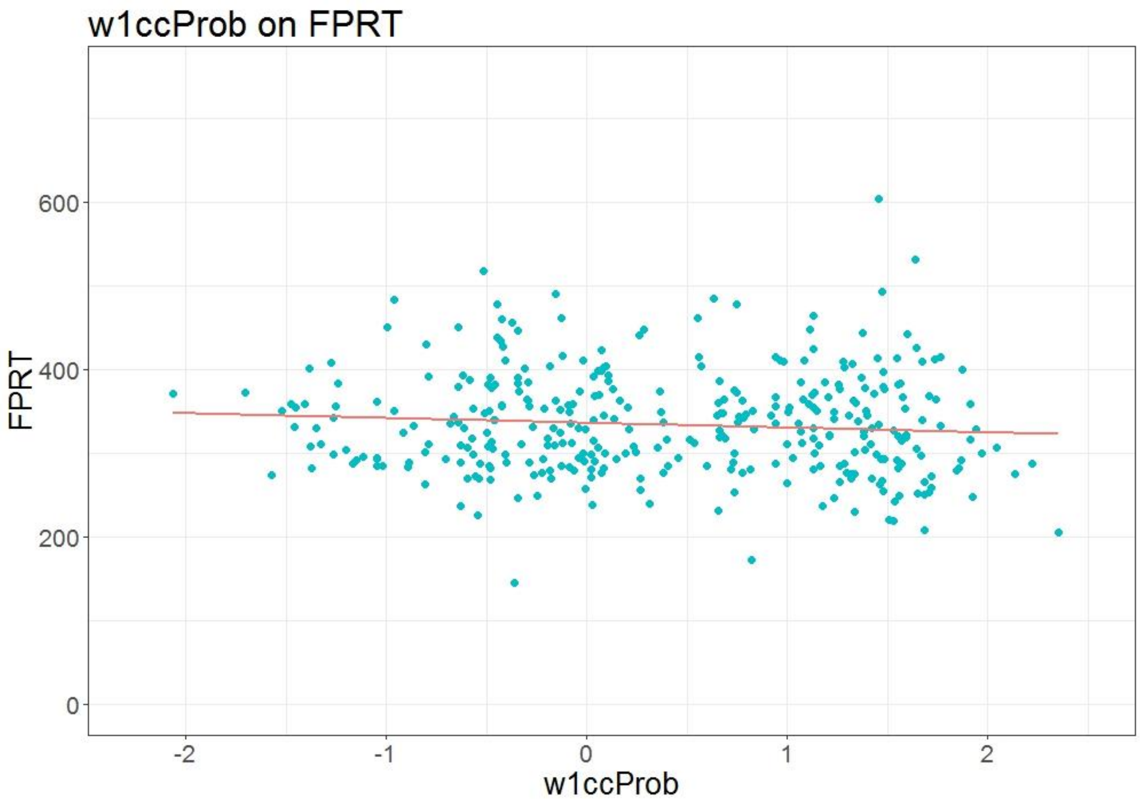


Figure 66: Effect of w2ccProb on reading time; first pass reading time decreases as w2ccProb increases.

An interaction was found between w2ccProb and w1ccProb, *β* = 6.522, *t*(4369) = 2.897, *p* = 0.004, likely due to w2ccProb decreasing reading speeds more sharply (as compared to w1ccProb) as it increases (Figure 67).



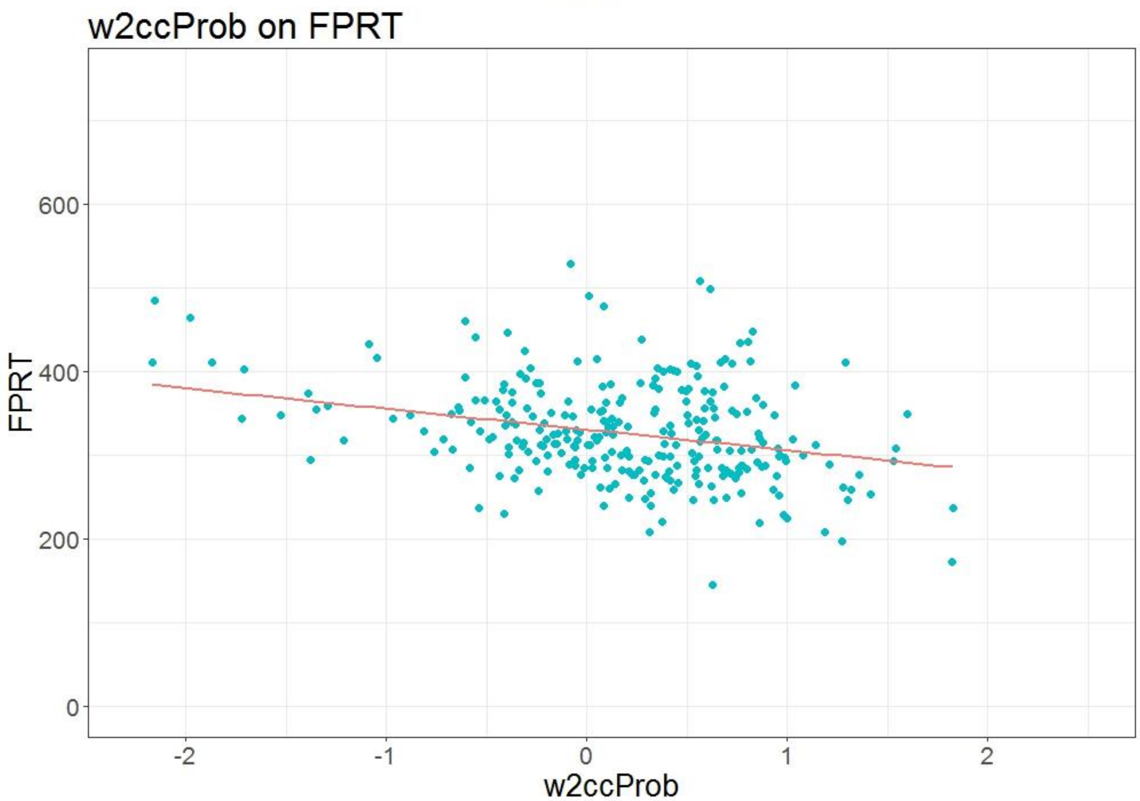


Figure 67: Comparison of effects of w1ccProb and w2ccProb on reading time. Reading speed decreases as both w1ccProb and w2ccProb increase, but it decreases at a faster rate as w2ccProb decreases.

No effect of w1ccProb was found, but a trend for an interaction between w1ccProb and language was found, $\beta = -5.557$, $t(4369) = -1.675$, $p = 0.10$ (Figure 68). The trend shows that first pass reading time tended to decrease as w1ccProb increased for both languages, with the effect more evident in Spanish.

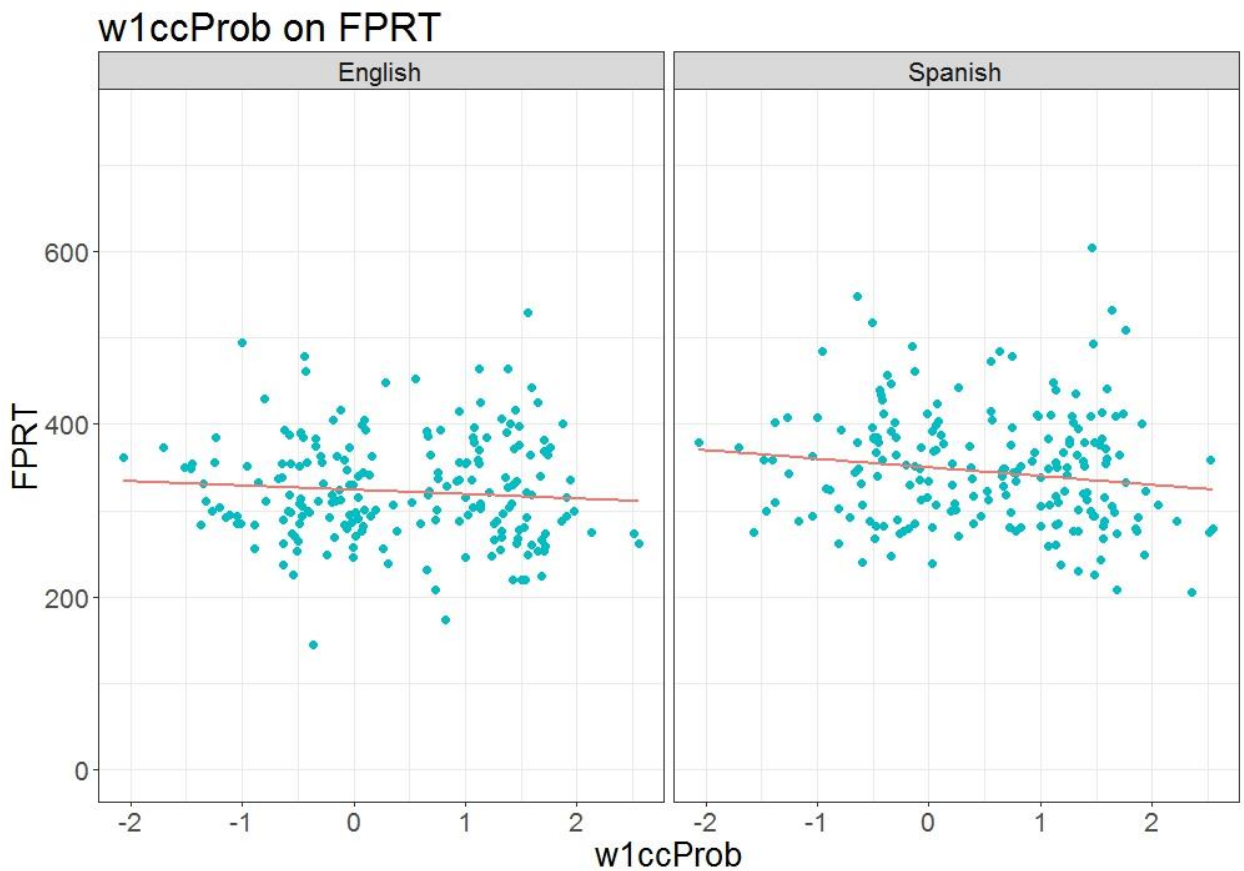


Figure 68: Effect of w1ccProb on first pass reading time separated by language; first pass reading time decreases as w1ccProb increases.

We also found a language effect, $\beta = 15.343$, $t(4369) = 2.487$, $p = 0.01$, with English words being read faster than Spanish words. This is likely due to the fact that most participants were educated in English, and used Spanish orally. Nouns in mixed language trials were read slower than in same language trials, $\beta = 12.645$, $t(4369) = 3.043$, $p = 0.003$, indicating a language switch cost. Though not significant, the large beta value for the interaction between language and congruency suggests that the language switch cost might be greater for Spanish-English trials than English-Spanish trials, $\beta = -11.319$, $t(4369) = -1.402$, $p = 0.16$. The pattern of results hints that this difference in switch cost seems to be primarily driven by Spanish-English word pairs with high w2ccProb values (most likely adv-N pairs) showing a greater grammatical predictability effect than similar high w2ccProb value English-Spanish word pairs, $\beta = -14.359$, $t(4369) = -1.388$, $p = 0.17$. The pattern only exists in relation to the w2ccProb values; a similar pattern was not found for trials with differing w1ccProb values in different language conditions, $\beta = 10.113$, $t(4369) = 1.276$, $p = 0.20$.

Figure 69 shows the 95% confidence interval for the main variables of interest, in milliseconds; the dotted line indicates the baseline (English, congruent, grammatical).

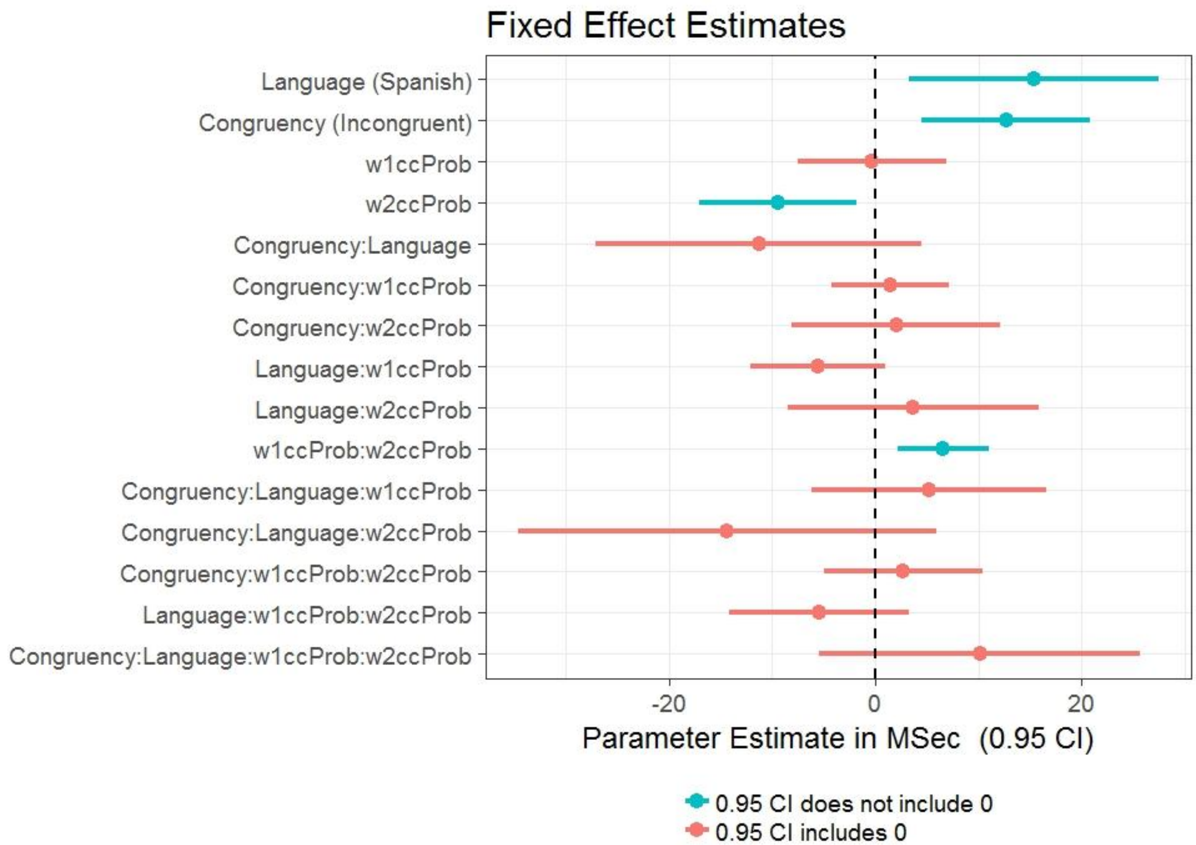


Figure 69: Fixed effects estimates for reading time in Experiment 3.

Of the control variables, we found main effects for the frequency and length of the second word, with more frequent and shorter words being read faster. (Frequency of Word 2: β = -16.464, $t(4369)$ = -3.491, p = 0.001; Length of Word 2: β = 46.577, $t(4369)$ = 11.786, $p < .001$) We found a spillover length effect, β = 9.353, $t(4369)$ = 3.622, $p < .001$, with shorter first words leading to faster reading times. No spillover frequency effect was found. Participants became faster as the experiment went on, β = -20.354, $t(4369)$ = -3.674, $p < .001$, and they were faster the second session, β = 9.836, $t(4369)$ = 2.234, p = 0.03, showing a practice effect.

Table 26 shows the results for the fixed variables of the model shown in Figure 65. *P*-values were generated using the lmerTest package (Kuznetsova et al., 2016).

Table 26: Results of the linear mixed-effects model for reading time in Experiment 3. Significant results are bolded.

Variable	*β*	*t*-value	*p*-value
Intercept	**336.287**	**29.201**	**<.001**
TrialNumber	**-20.354**	**-3.674**	**<.001**
SessionNumber	**9.836**	**2.234**	**0.03**
WordFrequency	**-16.464**	**-3.491**	**0.001**
WordLength	**46.577**	**11.786**	**<.001**
SpilloverFrequency	2.381	0.643	0.52
SpilloverLength	**9.353**	**3.622**	**<.001**
Language [Spanish]	**15.343**	**2.487**	**0.01**
Congruency [Incongruent]	**12.645**	**3.043**	**0.003**
w1ccProb	-0.350	-0.096	0.92
w2ccProb	**-9.429**	**-2.436**	**0.02**
Congruency*Language	-11.319	-1.402	0.16
w1ccProb*Congruency	1.449	0.501	0.62
w2ccProb*Congruency	2.003	0.389	0.70
w1ccProb*Language	-5.557	-1.675	0.10
w2ccProb*Language	3.663	0.588	0.56
w1ccProb*w2ccProb	**6.522**	**2.897**	**0.004**
Congruency*Language*w1ccProb	5.161	0.889	0.37
Congruency*Language*w2ccProb	-14.359	-1.388	0.17
Congruency*w1ccProb*w2ccProb	2.679	0.680	0.50
Language*w1ccProb*w2ccProb	-5.452	-1.230	0.22

Congruency*Language *w1ccProb*w2ccProb	10.113	1.276	0.20

Probability of an eye movement regression analysis. 7.45% of accurate trials included a regression from the second word back to the first. To see if these regressions affected the results of the reading time analysis, we looked at the probability that a trial included an eye movement regression back to the first string of characters. Table 27 shows the percent of trials with an eye movement regression as separated by language, language congruency, and grammaticality (as defined by category).

Table 27: Probability of an eye movement regression in Experiment 3.

	Exp. 3	
English-English	5.87%	
	det-N	5.96%
	adv-N	5.80%
Spanish-English	8.36%	
	det-N	11.07%
	adv-N	6.12%
English-Spanish	8.43%	
	det-N	9.67%
	adv-N	6.78%
Spanish-Spanish	7.23%	
	det-N	7.99%
	adv-N	6.29%
Same Language	6.52%	
	det-N	7.04%
	adv-N	6.01%
Mixed Language	8.40%	
	det-N	10.28%
	adv-N	6.42%
Word 2 (English)	7.07%	
	det-N	8.43%
	adv-N	5.95%
Word 2 (Spanish)	7.84%	
	det-N	8.86%
	adv-N	6.53%
ALL	7.45%	
	det-N	8.67%
	adv-N	6.21%

One pattern which is evident for all language conditions is that eye movement regressions are more common among grammatical trials as compared to ungrammatical trials. However, as we saw in Experiments 1 and 2, the different word properties of determiners and adverbs may be driving the pattern; it may not be the result of grammaticality.

To analyze the data, we performed a mixed-effects logistic regression using whether or not a participant made an eye movement regression as our dependent variable. Trials with a regression were considered hits, and trials without a regression were considered misses. The fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, session number, and our main variables of interest: language of the second word, language congruency, w1ccProb, and w2ccProb, as well as all interactions of those four variables. All variables were centered before being added to the model.

Random effects were taken into account for both item and subject. We again followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for trial number, length of the second word, length of the first word, frequency of the first word, w2ccProb, and language congruence; for items, slopes for w2ccProb, and language congruency, and trial number (Figure 70).

```
Model <- glmer(EyeRegression ~ TrialNumber  +
            SessionNumber +
            WordLength +
            LogWordFrequency +
            SpilloverLength +
            LogSpilloverFrequency +
            LanguageCongruence*WordLanguage*w1ccProb*w2ccProb +
            (0 + w2ccProb |item) +
              (0 + LanguageCongruence |item) +
              (0 + TrialNumber |item) +
            (1 |subject) +
              (0 + WordLength |subject) +
              (0 + SpilloverFrequency |subject) +
              (0 + SpilloverLength |subject) +
              (0 + w2ccProb |subject) +
              (0 + LanguageCongruence |subject) +
              (0 + TrialNumber |subject),
            data=bilingual, family= binomial(link = 'logit'),
            control=glmerControl(optimizer = 'bobyqa',
            optCtrl=list(maxfun=100000)))
```

Figure 70: Logistic mixed-effects model for probability of an eye movement regression in Experiment 3.

Neither w1ccProb nor w2ccProb are significant, despite the fact that if we look at grammaticality categorically, det-N pairs seem to have more regressions than adv-N trials. As in previous studies, the variance is likely caught by the spillover length, as length and grammaticality are correlated (between w1ccProb & word 1 length: $r(4401) = -0.75$, $p < .001$; between w2ccProb & word 1 length: $r(4401) = -0.37$, $p < .001$). An interaction was found between w1ccProb and language congruency, $\beta = 0.245$, $t(4370) = 2.013$, $p = 0.04$. Looking at the graph in Figure 71, we can see that the rate of eye movement regression seems to increase as w1ccProb increases (corresponding to higher grammaticality), with the effect being seen more in incongruent trials than in congruent trials.

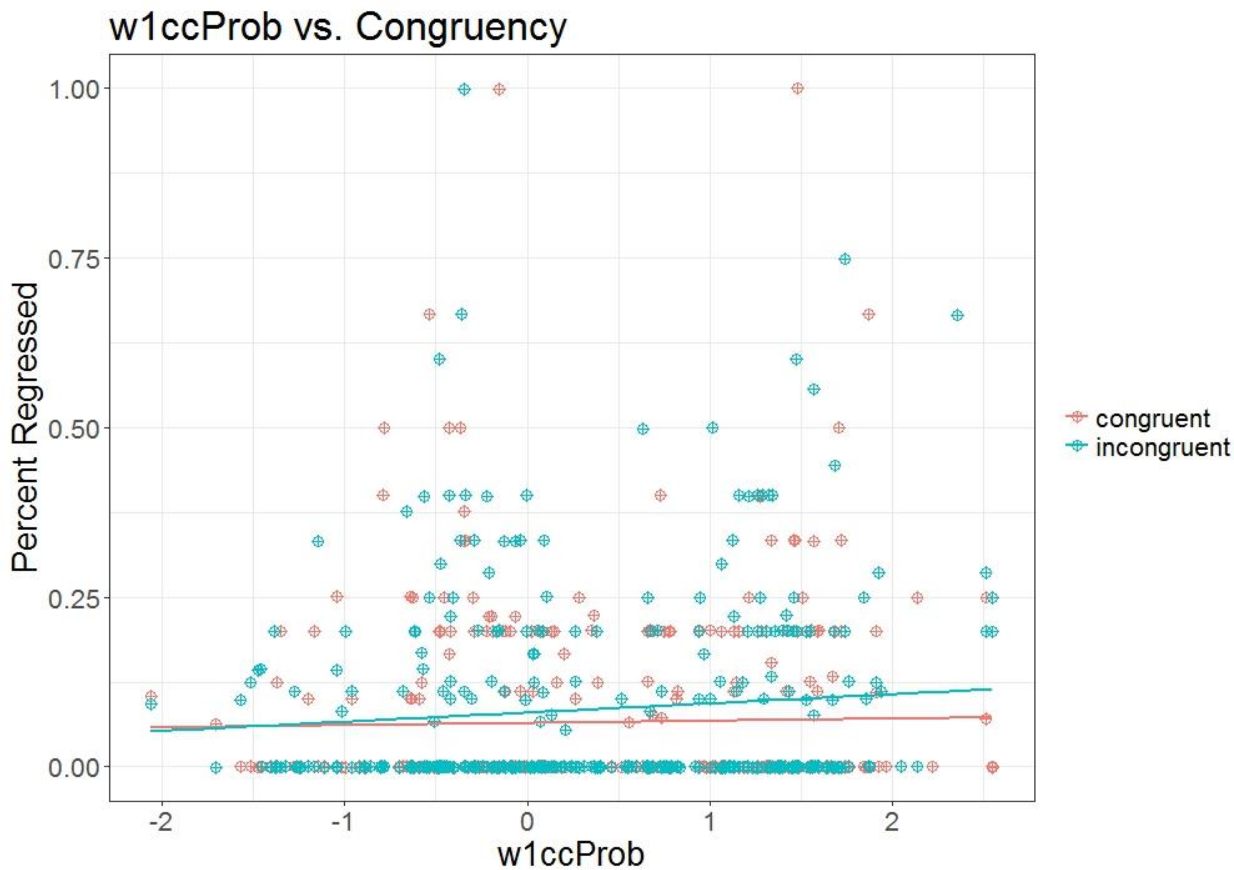


Figure 71: Congruency x w1ccProb interaction for probability of an eye movement regression in Experiment 3.

We found no main effect of language nor of congruency, but there was a trend for an interaction between the two variables, $β$ = -0.629, $t(4370)$ = -1.810, p = 0.07. There are similar rates of eye movement regression for the mixed language trials for both languages, but though the regression rates are less for the same language trials (~8.4%), less same language English trial have regressions (5.87%) than same language Spanish trials (7.23%).

A main effect of trial number was found, with participants being more likely to make a regression early in the experiment as compared to later in the experiment, $β$ = -0.725, $t(4370)$ = -4.649, $p < .001$. There was also a trend for participants to be more likely to make a regression in the first session as compared to the second session, $β$ = 0.243, $t(4370)$ = 1.771, p = 0.08. A main effect of length of Word 1 was found, with shorter Word 1 words being more likely to be regressed to, $β$ = -0.325, $t(4370)$ = -2.875, *p* = 0.004. A trend for the length of Word 2 was also found, with participants being more likely to regress the longer Word 2 is, $\beta = 0.178$, $t(4370) = 1.815$, $p = 0.07$, though it did not reach significance.

Reading time analysis for trials without an eye movement regression. No main effects of grammaticality were found in the probability of eye movement regression logistic regression. Since length of Word 1 and category are collinear, by removing trials with eye movement regressions we remove noise due to such regressions. We ran a linear mixed-effects regression using first pass reading time as the dependent variable only on trials without a regression.

As in the previous analyses, the fixed variables included length and log frequency of the second word, length and log frequency of the first word, trial number, session number, and our main variables of interest: language of the second word, language congruency, w1ccProb, and w1ccProb, as well as all interactions of those four variables. All variables were centered before being added to the model.

Random effects were taken into account for both item and subject. We followed a procedure similar to Barr et al., 2013 to remove terms and find the best estimators for the random effects, until obtaining a convergent, parsimonious model. The final model included for subjects, a random intercept, and slopes for session number, word length of the second word, frequency of the second word, language of the second word, and trial number; for items, a random intercept, and slopes for language congruency and trial number (Figure 72).

```
Model <- lmer( FPRT ~ TrialNumber  +
            SessionNumber +
            WordLength +
            LogWordFrequency +
            SpilloverLength +
            LogSpilloverFrequency +
            LanguageCongruence*WordLanguage*w1ccProb*w2ccProb +
            (1 |subject)
              (0 + SessionNumber |subject) +
              (0 + WordLength |subject) +
              (0 + LogWordFrequency |subject) +
              (0 + WordLanguage |subject) +
              (0 + TrialNumber |subject) +
            (1 |item) +
              (0 + LanguageCongruence |item) +
              (0 + TrialNumber |item),
            data=bilingual[no regression,], REML=F)
```

Figure 72: Linear mixed-effects model for reading time analysis excluding trials with an eye movement regression in Experiment 3.

The main reason to look at this model was to see if the main predictability effects of grammaticality would persist if the trials without regressions were analyzed alone. As with the model including all the trials (Figure 65), a main effect of w2ccProb is found, with participants becoming faster as it increases (β = -9.221, $t(4043) = -2.401$, $p = 0.02$). As in the model including eye movement regressions, the results indicate that Spanish-English trials had a greater cost than English-Spanish trials, β = -12.006, $t(4043) = -1.223$, $p = 0.22$, though this is not significant. No effect of w1ccProb was found, nor any interactions including w1ccProb.

Mixed language trials were read more slowly than same language trials, β = 16.520, $t(4043) = -4.533$, $p < .001$, showing a language switch cost. Trials with the second word in Spanish were read more slowly than those with the second word in English, β = 12.494, $t(4043) = 1.967$, $p = 0.05$. An interaction was found between congruency and language of the second word, showing differences in switch cost depending on whether the language of the second word was in English or Spanish, β = -16.899, $t(4043) = -2.210$, $p = 0.03$. In other words, though nouns were read faster in same language trials than mixed language trials, the same effect did not reach significance for trials in which the second word was in Spanish (Figure 73).

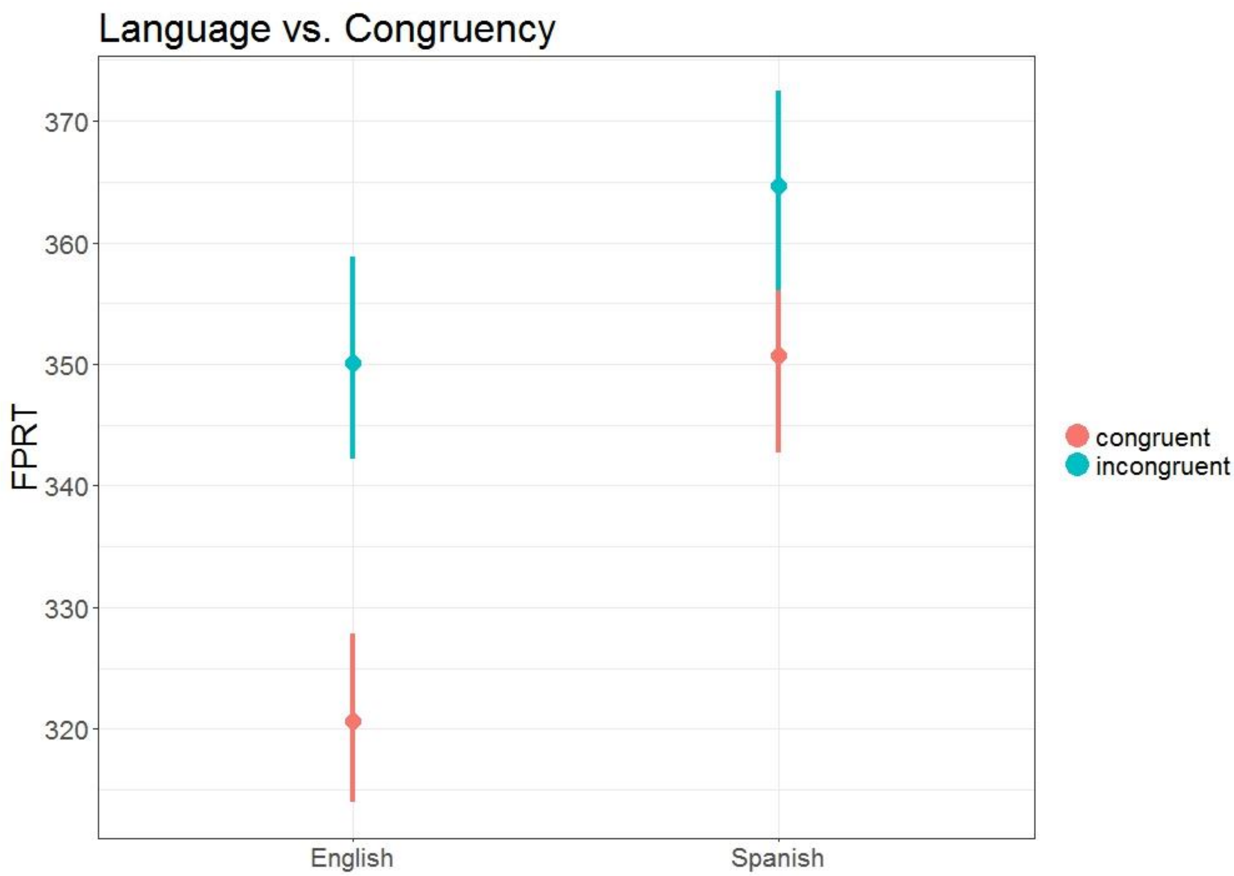


Figure 73: Congruency x language interaction for reading time analysis excluding trials with an eye movement in Experiment 3. Language refers to the language of the second word; error bars represent the standard error of the mean.

Of our control variables, a main effect of trial number and session were found, showing a practice effect, as participants became faster as the experiment went on (trial number: β = -28.151, $t(4043)$ = -5.194, $p < .001$), and were faster the second session as compared to the first (session number: β = 11.894, $t(4043)$ = 2.653, p = 0.01). Main effects of frequency and length were found for the second word, with more frequent words being read faster, β = -16.553, $t(4043)$ = -3.309, p = 0.001, and shorter words read faster, β = 49.844, $t(4043)$ = 11.474, $p < .001$. Table 28 shows the results for the fixed variables of the model shown in Figure 72.

Table 28: Results of the linear mixed-effects model excluding trials with an eye movement regression in Experiment 3. Significant results are bolded.

Variable	*β*	*t*-value	*p*-value
Intercept	**343.146**	**28.760**	**<.001**
TrialNumber	**-28.151**	**-5.194**	**<.001**
SessionNumber	**11.894**	**2.653**	**0.01**
WordFrequency	**-16.553**	**-3.309**	**0.001**
WordLength	**49.844**	**11.474**	**<.001**
SpilloverFrequency	1.925	0.536	0.59
SpilloverLength	3.162	1.270	0.20
Language [Spanish]	**12.494**	**1.967**	**0.05**
Congruency [Incongruent]	**16.520**	**4.533**	**<.001**
w1ccProb	-1.714	-0.487	0.63
w2ccProb	**-9.221**	**-2.401**	**0.02**
Congruency*Language	**-16.899**	**-2.210**	**0.03**
Congruency*w1ccProb	3.102	1.121	0.26
Congruency*w2ccProb	5.430	1.111	0.27
Language*w1ccProb	-4.607	-1.418	0.16
Language*w2ccProb	-0.162	-0.026	0.98
w1ccProb*w2ccProb	2.843	1.268	0.20
Congruency*Language*w1ccProb	5.077	0.910	0.36
Congruency*Language*w2ccProb	-12.006	-1.223	0.22
Congruency*w1ccProb*w2ccProb	2.440	0.643	0.52
Language*w1ccProb*w2ccProb	-2.513	-0.578	0.56

Congruency*Language*w1ccProb*w2ccProb	8.253	1.079	0.28

Discussion

Experiment 3 focused on taking into account the fact that words are ambiguous in regards to category, and the implications of that fact, especially as to how it relates to predictability effects. First, importantly, we were able to account for categorical ambiguity in words by using conditional probabilities based on the grammatical categories a word could be, and the probability of that category either preceding or following the other word of the word pair. This resulted in two values, one being the probability of the category of the second word given the first word (w1ccProb), and the other the probability of the second word given the category of the first word (w2ccProb). Notably, this allowed us to discard the abstract binary 'grammaticality' variable used in Experiments 1 and 2, replacing it with two variables, capturing a more fine-grained, concrete relationship between words.

As in Experiments 1 and 2, our primary analysis was the linear mixed-effects model looking at the reading speed of the noun (in this experiment being defined as a word that appears most often as a noun in language). We were interested in seeing whether there was a relationship between the reading speed of the noun and the conditional probabilities. As higher conditional probability values means that a noun most often follows the category of the first word (w2ccProb, analogous also to lower levels of surprisal) or that a determiner most often precedes the category of the second word (w1ccProb), our predictions were similar in regards to the conditional probabilities. To put it simply, faster reading speeds were expected to be associated with higher conditional probability values. As we had no specific hypothesis as to which (or whether both) conditional probability is associated with predictability effects, a positive relationship between either conditional probability value and reading speed would be taken to be indicative of grammatical predictability effects.

We found a main effect of w2ccProb, with participants reading the noun faster as the w2ccProb increased. This is in line with our hypothesis, suggesting that grammatical predictability effects are being captured by w2ccProb. It did not interact with language, showing that participants showed predictability effects regardless of the language of the second word. Additionally, no interaction was found between w2ccProb and language congruency, showing that the effect carries across languages.

Though we found a main effect of w2ccProb, no similar main effect of w1ccProb was found. Nevertheless, a strong trend between language and w1ccProb was found. Participants tended to read the noun faster as w1ccProb increased whatever the language of the noun, though the effect was more evident when the noun was in Spanish. A possible w1ccProb effect may be captured by the interaction, resulting in no main effect of w1ccProb. As the interaction did not reach significance, it is impossible to draw more definitive results from the data, though it leaves the question open of whether w1ccProb affects predictability.

One piece of evidence suggesting that both w1ccProb and w2ccProb have effects on predictability is the accuracy analysis, which found main effects for both conditional probabilities; as the value of each of the conditional probability increased, so too did the accuracy. Unfortunately, the high degree of accuracy shown by the participants resulted in ill-fitting logistic regression models; instead, a linear regression was performed using average values per language condition and grammaticality by participant. The mean conditional probability values for stimuli may heighten effects, as it results in less overlap in grammatical and ungrammatical stimuli for both conditional probabilities.

Limitations

We are interested in how syntax is affecting reading speed, which leads to a concern is that semantics may be in part driving the effect. To prevent semantic priming, especially in the case of ungrammatical word pairs, we specifically made pairs which are not logically linked semantically. For example, '*emphatically machine*' could appear as a trial, but not '*emphatically shout*'. This semantic distancing was not possible in many grammatical word pairs given the nature of determiners. As a result, ungrammatical word pairs may have had longer processing times, as the most common categorical reading referred to impossible/unlikely semantic concepts; in contrast, the grammatical word pairs did not.

Another concern is that we did not perform any linguistic assessments on our participants, instead only asking participants to self-rate their English and Spanish abilities. Such a measure is not objective, and scores varied among participants as they used different rating criteria. No further assessments were performed due to the length of the study (2 one-hour sessions which sometimes overran). Linguistic proficiency has been shown to have an effect on reading times, however; in most of the analyses for Experiments 1 and 2, there was an effect of Spanish proficiency, and in some English proficiency was also found to be significant. By not including language assessments we are assuming that participants are equally strong in English and Spanish; the data from the current experiment would suggest otherwise, since participants read English words faster than Spanish words. Not including language proficiency adds noise to the data.

Conclusions and Implications

In summary, our results accounted for categorical ambiguity and found a main effect of w2ccProb, finding evidence for a granular grammaticality effect. As with binary grammaticality, the effect of w2ccProb was found regardless of the language of the words in the word pair, and there were no interactions with language nor with language congruency. Taken together, this suggests that predictability due to w2ccProb is language-independent.

Whether or not w1ccProb is having an effect on predictability is inconclusive, as no main effect was found, but a trend between w1ccProb and language suggests that it might be having an effect. Not only does the trend suggest that w1ccProb may play a role in regards to predictability, but it also suggests it may be more sensitive to language, unlike w2ccProb. W1ccProb hints that the lexical identity of the first word may be playing a role in predictability effects, though how exactly is unclear.

Our results are similar to those found in Experiments 1 and 2, suggesting that the conditional probability variables encompass grammaticality. Since we did not measure any participant-specific traits, such as language proficiency or language use, we can not infer as to how they may interact with the conditional probability values. Further research must be conducted to see if this is the case.

Chapter 6

General Discussion

Findings

Over half of the world is a bilingual or language learner; monolingualism is the minority, not the norm. Much of language research, however, has been performed on monolinguals, for several reasons, both experimental and practical. For example, it lowers the number of variables and allows us to examine how language works in general. A monolingual's language tends to be more uniform than a bilingual's; for example, children born to English-speaking households, educated in English, and in college are likely to have similar linguistic skills. Bilingualism, in contrast, is more varied, due to the fact that a person will have more than one language to use, did not necessarily learn the languages at the same time, and may use the languages in different ways, such as in the home, at school, or at work.

One of the issues concerning language use is code-switching, when a speaker of more than one language switches from one language to the other in the middle of a conversation, or sentence. Code-switching is of particular interest because it is a point where a bilingual is using both their languages concurrently. We are particularly intrigued by code-switches that occur mid-sentence, e.g., intrasentential code-switches, and what these code-switches can tell us about the nature of what is shared between languages.

This dissertation focused on grammatical predictability effects. Our main question of interest was to find whether or not grammatical predictability persisted across languages. To look at this question, we performed three experiments, each further delving into the question. In our summary of results, we shall focus on the reading time analyses, as they were the most consistent. In Experiment 1, we looked at grammatical predictability in Spanish-English bilinguals, assuming that grammaticality is a binary factor—a sentence is or is not grammatical. We found that bilinguals showed reliable predictability effects, and that the magnitude of the effects in English and Spanish did not differ depending on language.

Most interestingly, though bilinguals showed a language switch cost (equal regardless of English-Spanish or Spanish-English), the grammatical predictability effect did not interact with language congruency, i.e., whether or not there had been a language switch. Taken together, these results indicate that grammatical predictability effects in bilinguals persist across language. In fact, the effect size of grammatical predictability tends to be larger, not smaller in mixed language trials. Otherwise stated, grammatical predictability is language independent.

The results are most supportive of the shared syntax account of bilingual language representation (Hartsuiker & Pickering, 2008), which states that language is a property of words, not necessarily of grammatical structures, at least for structures which are shared across languages. Words populate the structures regardless of their language, and so effects due to grammar (such as grammatical predictability) are found regardless of the language of the words which fill the structure. The fact that we found a main effect of grammatical predictability that did not interact with either language congruency or language is easily explained by this model. Results did not support a view of (mostly) separate bilingual language representation, as is presented in the bilingual production model (de Bot, 1992), since the grammatical predictability effect was just as large for mixed language trials as same language trials.

In Experiment 2, we looked at how linguistic proficiency interacted with grammatical predictability, specifically looking at advanced L2 learners of Spanish. Looking at L2 language learners allowed us to focus on how proficiency affects grammatical predictability since we were able to look at a population who showed high variability in Spanish linguistic knowledge (after having reached a particular proficiency) but showed low variability in English linguistic knowledge (since they were experts in the language).

A strong trend was found for grammatical predictability in language learners. Similar to the bilingual results, grammatical predictability did not interact with either language or language congruency. This finding is despite the fact that an effect of Spanish proficiency was found (participants were faster to read the noun the more proficient they were). Further evidence that language learners show grammatical predictability effects comes from the combined data model. Looking at all the data together, we found an effect of grammatical predictability, but no effect of participant type, showing that both language learners and bilinguals display grammatical predictability. Overall, there was no difference in behavior between language learners and bilinguals, other than a strong trend of language learners being slower to read words in Spanish.

To make sure we were not overlooking proficiency effects, we checked to see if an interaction would be found between grammatical predictability and Spanish proficiency. No interaction was found between Spanish proficiency and language congruency, nor between Spanish proficiency and language congruency. This finding was a bit surprising, as it would seem logical for grammatical predictability to differ in regards to language knowledge; however, participants showed similar grammatical predictability effects in both English and Spanish. Moreover, as with bilinguals, language learners showed similar grammatical predictability effects regardless of whether or not there was a language switch, signifying that grammatical predictability is language-independent.

The results support the shared syntax account, as do the bilingual findings, though language learners showed slightly lesser grammatical predictability in mixed language trials. This could indicate that the language systems are not fully independent, which would not invalidate the bilingual production model. One question which Experiment 2 fails to answer is whether or not the methodology provides support for the Revised Hierarchical Model (Kroll & Stewart, 1994). The model states that beginning language learners translate L2 words into their L1 automatically to access word meanings; in other words, there is no direct link between semantic information and L2 word forms. If this is the case, then (L2, L1) word pairs should show a grammatical predictability effect similar to the (L1, L1) word pairs, since the automatic translation occurs at the first word, and we are looking at effects only on the second word. We would see a language switch cost in (L1, L2) word pairs due to translation, and it is not certain whether grammatical predictability effects would remain past the translation. To simplify, according to the RHM, we would not expect to see the same pattern of results in both types of mixed language pairs.

As grammatical predictability did not interact with language congruency, it suggests that as in the bilinguals, grammatical predictability occurs regardless of the language of the word pairs (at least in cases of shared grammaticality). This does not support the RHM; however, as the RHM states that automatic translation is a transitory stage before forming connections between L2 word forms and conceptual representations, it also does not provide evidence against it. Our participants were advanced learners, who may have already formed direct connections between L2 word forms and the semantic representation corresponding to those word forms.

Having tested both language learners and bilinguals meant that we had participants whose lifestyles would lead them to use language differently. Language learners acquire Spanish in school, and they have used it primarily in Spanish classes. Though some participants had spent some time in Spanish-speaking countries, they used English primarily for education, were living in an English-speaking country, and reported speaking English at home. Bilingual participants tended to have more opportunities to speak Spanish, either because they used Spanish primarily at home, had more Spanish-speaking friends, or could be classified as an English L2 learner. Otherwise stated, more opportunities to speak Spanish with Spanish-English bilinguals can be thought of as more opportunities to code-switch.

As we previously mentioned, no differences were found in grammatical predictability between language learners and bilinguals, in terms of their reading times. The main difference was in regards to reading Spanish, though it did not reach significance. As the model not including participant type better suited the data, it suggests that difference in performance was captured by proficiency. Linguistic ability in general might be affecting grammaticality, but both language learners and bilinguals were sufficiently proficient in both English and Spanish to show indistinguishable behavior, at least in regards to grammatical predictability.

Since we did not find a difference between language learners and bilinguals, we looked at more specific behavioral variables which might affect grammatical predictability. It could be the case that a variable that might seem to be logically highly correlated with participant type is not, and grammatical predictability is instead affected by that variable; if so, we would not see a difference between groups, but a behavioral difference would be shown to affect grammatical predictability. The first variable we looked at was estimated percentage of time spent code-switching with top conversational partners, as this was meant to capture a rough estimate of a participant's average language use. Though as in previous models a main effect of grammatical predictability is found (~12msec), this was not affected by estimated percentage of time spent code-switching with top conversational partners, despite the fact that participants showed a wide range of time spent code-switching.

We also looked at participants' rating of ungrammatical sentences on the code-switching sentences assessment task. Linguistic proficiency appears to be the most important factor affecting grammatical predictability; participants' rating of ungrammatical code-switching sentences (UCS) are a measure of code-switching proficiency, and could logically affect the results. UCS ratings were not found to affect grammatical predictability; however, an interaction was found between UCS rating and language congruency, with the participants who rated the sentences lowest showing the least language switch cost. This is due to slower reading speeds on same language trials for participants who rated the UCS lower as compared to participants who rated the UCS higher. This might be indicative of participants' views on code-switching, especially bilinguals who view code-switching as being undesirable; however, more targeted assessments would have to be done to specifically test for how personal differences in code-switching beliefs might affect the results. The pattern might not even be indicative of code-switching viewpoints; instead, it may signify individual differences in grammaticality perception; however, as it stands now, no conclusions can be drawn from this interaction between UCS rating and language switch cost.

The methodology used in Experiments 1 and 2 was effective in that it allowed us to find a grammatical predictability effect, and to look at participant factors that may affect it. To do so, stimuli were carefully chosen to have high frequency words all participants would have long time knowledge of, and so could easily recognize. In addition, we attempted to minimize grammatical category ambiguity for the words chosen; however, this was not completely eliminated, as many words (in both English and Spanish) can have some level categorical ambiguity. Nevertheless, any remaining categorical ambiguity was treated as unavoidable noise, and all word pairs were classified as either 'grammatical' or 'ungrammatical', rather than falling into a possible spectrum.

Unlike the first two studies which focused more on individual differences in participants, Experiment 3 focused more on the stimuli used in the study. Specifically, the assumption that words had no categorical ambiguity and word pairs had no categorical ambiguity was questioned. Rather than look at a grammatical predictability effect, we instead looked at two variables: the probability of word one given the category of word two (w1ccProb), and the probability of word two given the category of word one (w2ccProb). Categorical ambiguity was treated in that 'category of word' was a sum consisting of the proportion of times that word appeared as that category. Doing so allowed us allowed us to calculate fine-grained, continuous predictors indicative of grammaticality.

We found an effect of w2ccProb, in that words with higher w2ccProb values had faster reading times. This is analogous to measuring surprisal, as it is indicative of how likely the second word is to occur (Levy, 2008). Similar to the grammaticality effect in Experiments 1 and 2, w2ccProb is language-independent in that the same pattern is seen regardless of whether or not there is a language switch.

Though we did not find a main effect of w1ccProb, a trend between language and w1ccProb suggests that it may be having an effect, even if it did not reach significance. As with w2ccProb, reading time decreases as w1ccProb increases, suggesting that grammatical predictability is sensitive to the lexical identity of the first word. The effect is present in both same language and mixed language word pairs; however, a greater decrease can be seen when the second word is in Spanish.

Not having testing linguistic proficiency in Experiment 3, which was shown to be significant in Experiments 1 and 2, may indicate that there is increased relative noise in the results, at least in comparison to Experiments 1 and 2. More research would have to be performed to draw stronger results, and to see how proficiency affects w1ccProb and w2ccProb; presently, the main finding is that participants show language-independent w2ccProb.

Future Directions

In the series of experiments discussed in this dissertation, we adapted an existing task (the lexical decision task) in a novel way to look at whether or not grammatical predictability effects are found across languages. We have found a pattern which we take to be evidence for grammatical predictability—faster reading times for nouns following determiners as compared to nouns following adverbs when grammaticality is binary, and faster reading times as conditional probabilities increase. Furthermore, the effect is language-independent, found whether or not there is a language switch, and whatever the language of the words. The effect has been taken to be support of the shared syntax account of how more than one language can be represented. Moreover, the same pattern of results is found in both language learners and bilinguals, using different stimuli, and whether or not grammaticality is encompassed as a binary variable or replaced with conditional probabilities.

Many questions yet remain, however. One important question that is not altogether answered is how linguistic proficiency is affects grammatical predictability, the main question we addressed in Experiment 2. The participants tested in the study were all advanced language learners; though they scored lower than bilinguals in Spanish proficiency, all participants had studied Spanish for at least 5 years, and had taken multiple courses taught in Spanish; in other words, all participants were carefully selected to have high command of Spanish and be experts in English. As mentioned previously, the advanced language learners performed very similarly to bilinguals, with the main difference between the groups seen in the fact that language learners were slower to read in Spanish than bilinguals.

Looking at less proficient language learners would give us a greater difference between native and second language proficiency. Since the stimuli in Experiment 2 were selected to as to have been among the vocabulary presented to students in their first year, language learners who had not yet completed the language sequence could be tested. This would also allow us to take a closer look at the Revised Hierarchical Model. The participants in Experiment 2 are likely to have already formed connections between Spanish word forms and word concepts; participants with low proficiency are more likely to not have formed the connections, and instead depend on automatic translation before being able to access the word concept before being able to recognize whether or not a string is a Spanish word, which would lead to differences in in the two types of mixed language pairs. Furthermore, lower linguistic proficiency may lead to grammatical predictability being found only in English. It would be worthwhile to test directly, especially since beginning students are more likely to be forced to code-switch due to low vocabulary in Spanish, and so would be modeling behavior seen in the classroom.

Another question remaining is whether the same conditional probability predictability patterns would be replicated with different stimuli. In Experiment 3 we found evidence of language-independent w2ccProb, suggesting a grammatical predictability effect. It would be beneficial to analyze the data in Experiment 1, replacing grammaticality with the conditional probability variables, to see if we also find language-independent w2ccProb using different stimuli and a different group of bilinguals.

Analyzing the data in such a way would also allow us to look at whether participant differences, as collected in the language assessments, has an effect. In Experiment 1, we found an effect of Spanish proficiency in every analysis, and some analyses also showed an effect of English proficiency. As we did not collect such data in Experiment 3, differences in proficiency might be adding noise to the data.

It would also be beneficial to analyze the data in Experiment 2 in the same manner, to see if language learners show the same predictably patterns as bilinguals in regards to conditional probabilities. Experiments 1 and 2 showed that bilinguals and language learners showed the same patterns when grammaticality was treated as a binary variable; this might not necessarily carry over to conditional probabilities. Unlike grammaticality, which is based on the grammatical rules of language, conditional probabilities might be more sensitive to usage, as they are based on word frequency. Differences between bilinguals and language learners are more likely to appear, because although both groups use the same languages (which have the same grammatical rules), their usage patterns are different. Similarly, effects of code-switching use (which were not significant when using grammaticality as a variable) might also be found in conditional probability analyses.

The experiments performed looked only at cases where grammaticality was shared between languages. There were several reasons for doing so; natural Spanish-English code-switches are found only in cases of shared grammaticality (Poplack, 1980), and even then, not all code-switches are equally acceptable to bilinguals (Belazi et al., 1994; Toribio, 2001a). As we wished to focus on natural code-switches, and maximize the probability of grammatical predictably, we looked at word pairs that were equally grammatical in both English and Spanish.

Not surprisingly however, since English and Spanish are not the same language with different word forms, there exist word combinations in one language which are grammatical which are not grammatical in the other. For example, English adjectives always precede the noun; Spanish adjectives usually follow the noun, with a few exceptions (Zagona, 2002). In this case, grammaticality would be different in English and Spanish, and it can be thought of as being intermediate in mixed language pairs. As grammaticality is intermediate, we would not be able to treat it as an absolute, as we did in Experiments 1 and 2; however, conditional probabilities allow for a range of grammaticality.

It would also be advantageous to look at different languages. We looked at Spanish-English word pairs, hoping to extrapolate about language representation. However, different bilinguals of different languages do not necessarily follow the same rules. Spanish-English code-switches only occur in cases of shared grammaticality, but this is not the case in other pairs of languages. Natural code-switches in different languages sometimes occur when the code-switch is grammatical in only one language, as in Welsh-English code-switching (Deuchar, 2005), or Spanish-Hebrew code-switching (Berk-Seligson, 1986). In some observed cases, code-switching leads to ungrammatical forms in both languages, as in French-Moroccan Arabic code-switches (Bentahila & Davies, 1992).

In all the cases that lead to code-switching occurring in places that grammaticality is not shared, the languages are more different than Spanish and English. It would be beneficial to look at languages that differ more grammaticality, as this may lead to more insights on how languages are represented. The shared syntax account is a good model for Spanish and English, because in many cases grammaticality is shared, and code-switches seem the natural result. In other words, Spanish-English code-switches occur in such a case because the similar grammar allows it, and so the speakers do not need to form a separate code-switching grammar. However, since code-switching is seen regardless of the similarity of the languages, it may be the case that different languages are not necessarily represented in the same way, and predictability would be more influenced by experience. It would be advantageous to look at a language pair consisting of two languages with highly different grammars, such as Spanish-Basque, which is formed of languages from different families but that share a similar writing system.

Conclusion

Over half of the people of the world are bilingual, and as immigration and globalization continue increasing, it is likely that number will increase. Many questions still remain as to how bilingualism works however, and how languages affect each other in a speaker. Code-switching is at the forefront of language interaction, as it is when both languages are currently activated in a speaker. As such, code-switching is a fruitful phenomenon to study, and can tell us much about the mechanics of bilingualism.

In this dissertation, we provided an in-depth investigation into grammatical predictability in Spanish-English bilinguals and language learners. Our findings have strong implications about how languages are represented and interact. The findings also have potential implications on the use of code-switching during language learning, as well as what aspects of experience most affect the processing of code-switches.

APPENDIX A:

lexTALE-ESP

Hi, this is a test of Spanish vocabulary. On the next page you will find approximately 90 sequences of letters that look "Spanish". Only some of them are real words. Please, indicate the words you know (or of which you are convinced they are Spanish words, even though you would not be able to give their precise meaning). Be careful, however: Errors are penalised. So, there is no point in trying to increase your score by adding tallies to "words" you've never seen before! You have as much time as you like for each decision. This part of the experiment will take about 5 minutes.

	YES	NO
calpar	YES	NO
joten	YES	NO
sacapuntas	YES	NO
terzo	YES	NO
pellizcar	YES	NO
pulmones	YES	NO
batillón	YES	NO
zapato	YES	NO
tergiversar	YES	NO
pésimo	YES	NO
cadeña	YES	NO
hacha	YES	NO
antar	YES	NO
cenefa	YES	NO
asesinato	YES	NO
helar	YES	NO
yunque	YES	NO
regar	YES	NO
abracer	YES	NO
floroso	YES	NO
arsa	YES	NO
brecedad	YES	NO
ávido	YES	NO
capillo	YES	NO
lacayo	YES	NO

lampera	YES	NO
látigo	YES	NO
bisagra	YES	NO
secuestro	YES	NO
acutación	YES	NO
merodear	YES	NO
decar	YES	NO
alardio	YES	NO
pandilla	YES	NO
fatacidad	YES	NO
pauca	YES	NO
aviso	YES	NO
rompido	YES	NO
loro	YES	NO
granuja	YES	NO
estornudar	YES	NO
torpe	YES	NO
alfombra	YES	NO
rebuscar	YES	NO
cadallo	YES	NO
canela	YES	NO
cuchara	YES	NO
jilguero	YES	NO
martillo	YES	NO
cartinar	YES	NO
ladrón	YES	NO
ganar	YES	NO
flamida	YES	NO
candado	YES	NO
camisa	YES	NO
vegada	YES	NO
fomentar	YES	NO
nevar	YES	NO
musgo	YES	NO
tacaño	YES	NO
plaudir	YES	NO
besar	YES	NO
matar	YES	NO
seda	YES	NO
flaco	YES	NO
esposante	YES	NO
orgulloso	YES	NO
bizcocho	YES	NO
hacido	YES	NO
cabello	YES	NO
alegre	YES	NO

engatusar	YES	NO
temblo	YES	NO
polvoriento	YES	NO
pemición	YES	NO
hervidor	YES	NO
cintro	YES	NO
yacer	YES	NO
atar	YES	NO
tiburón	YES	NO
frondoso	YES	NO
tropaje	YES	NO
hormiga	YES	NO
pozo	YES	NO
empirador	YES	NO
guante	YES	NO
escuto	YES	NO
laúd	YES	NO
barato	YES	NO
grodo	YES	NO
acantilado	YES	NO
prisa	YES	NO
clavel	YES	NO

APPENDIX B:

lexTALE

This test consists of about 60 trials, in each of which you will see a string of letters. Your task is to decide whether this is an existing English word or not. If you think it is an existing English word, you click on "yes", and if you think it is not an existing English word, you click on "no".

If you are sure that the word exists, even though you don't know its exact meaning, you may still respond "yes". But if you are not sure if it is an existing word, you should respond "no".

In this experiment, we use British English rather than American English spelling. For example: "realise" instead of "realize"; "colour" instead of "color", and so on. Please don't let this confuse you.

This experiment is not about detecting such subtle spelling differences anyway.

You have as much time as you like for each decision. This part of the experiment will take about 5 minutes.

If everything is clear, you can now start the experiment.

	YES	NO
platery	YES	NO
denial	YES	NO
generic	YES	NO
mensible	YES	NO
scornful	YES	NO
stoutly	YES	NO
ablaze	YES	NO
kermshaw	YES	NO
moonlit	YES	NO
lofty	YES	NO
hurricane	YES	NO
flaw	YES	NO
alberation	YES	NO
unkempt	YES	NO
breeding	YES	NO

festivity	YES	NO
screech	YES	NO
savoury	YES	NO
plaudate	YES	NO
shin	YES	NO
fluid	YES	NO
spaunch	YES	NO
allied	YES	NO
slain	YES	NO
recipient	YES	NO
exprate	YES	NO
eloquence	YES	NO
cleanliness	YES	NO
dispatch	YES	NO
rebondicate	YES	NO
ingenious	YES	NO
bewitch	YES	NO
skave	YES	NO
plaintively	YES	NO
kilp	YES	NO
interfate	YES	NO
hasty	YES	NO
lengthy	YES	NO
fray	YES	NO
crumper	YES	NO
upkeep	YES	NO
majestic	YES	NO
magrity	YES	NO
nourishment	YES	NO
abergy	YES	NO
proom	YES	NO
turmoil	YES	NO
carbohydrate	YES	NO
scholar	YES	NO
turtle	YES	NO
fellick	YES	NO
destription	YES	NO
cylinder	YES	NO
censorship	YES	NO
celestial	YES	NO
rascal	YES	NO
purrage	YES	NO

pulsh	YES	NO
muddy	YES	NO
quirty	YES	NO
pudour	YES	NO
listless	YES	NO
wrought	YES	NO

APPENDIX C:

SPANISH-ENGLISH CODE-SWICHING SENTENCE JUDGMENT TASK

In this task, you will read different sentences that include a change of language. Please rate on a scale of 1 to 7 how acceptable the sentence sounds to you, with 1 meaning it sounds completely unacceptable and 7 meaning it sounds completely acceptable. There are no right or wrong answers, so judge the acceptability depending on your reaction to each sentence.

Toda mi familia came to visit.

(unacceptable) 1 2 3 4 5 6 7 (acceptable)

My friend and I have visto esa película.

(unacceptable) 1 2 3 4 5 6 7 (acceptable)

El médico le dijo que to lose weight would be good.

(unacceptable) 1 2 3 4 5 6 7 (acceptable)

I parked my car en frente del arból de manzana.

(unacceptable) 1 2 3 4 5 6 7 (acceptable)

Todos mis amigos who live nearby came over.

(unacceptable) 1 2 3 4 5 6 7 (acceptable)

El estudiante puede study during the train ride to campus.

(unacceptable) 1 2 3 4 5 6 7 (acceptable)

The detective deduced que el marido mató a su esposa.

(unacceptable) 1 2 3 4 5 6 7 (acceptable)

My parents did not visitar el año pasado.

(unacceptable)	1	2	3	4	5	6	7	(acceptable)

La autora escribió y published a new book last year.

(unacceptable)	1	2	3	4	5	6	7	(acceptable)

The children asked for jugo y algo para comer.

(unacceptable)	1	2	3	4	5	6	7	(acceptable)

APPENDIX D:

ACSES

INSTRUCTIONS: Place a '1' only in the light pink box that corresponds to your answer. (For instance, for question #1, if your first language is Spanish, place a '1' in the light pink box to the right of Spanish and leave the other pink boxes for that question blank.) Blue boxes require fill in the blank answers (words or numbers). Some questions have multiple questions each written on a separate row. **Please be sure to answer each of these multiple questions.**

Please do not go back and change answers after you complete a section.

SECTION 1

1.	**What is your first language?**	
	Spanish	
	English	
	Both Spanish and English	
	Other	
	If other, please specify:	

2.	**What is your second language?**	
	Spanish	
	English	
	Both Spanish and English	
	Other	
	If other, please specify:	

3.	**Do you speak any other languages?**	
	Yes	
	No	
	If yes, please specify:	

4. **Compared to English and Spanish, how often do you speak *other* languages?**

Rarely	
Sometimes	
Frequently	

For the rest of this survey, please respond for Spanish and English only.

In this survey, language switching is using two languages in the same conversation. Language switching and mixing mean the same thing.

might speak several sentences in English. Y luego cambio a español para acabar la conversación.

5. **Please rate on a scale of 1-7 (1 is very poor, 7 is very good) your ability in each language for the following categories:**

	Speaking	Comprehension	Reading	Writing
Spanish				
English				

SECTION 2

Part 1

6. **Which of the following best describes you?**

I have never mixed Spanish and English.	
I used to mix Spanish and English, but do not do it anymore.	
I never used to mix Spanish and English, but have recently started.	
I have mixed Spanish and English my whole life.	

For each of these questions, you answers should add up to 100%

Note: In this survey, anyone who speaks some Spanish and some English is a bilingual.

7. **Of the time you spend speaking to Spanish/English bilinguals, how much do you spend...**

Mixing your languages?		
Speaking one language at a time?		
	Total:	0

8. **Of the time you spend to speaking to Spanish English bilinguals, how much do you spend speaking to people who:**

Like to mix English and Spanish?		
Don't like to mix English and Spanish?		
	Total:	0

9. **Do you switch between languages within a conversation when speaking with...**

	Never	Rarely	Occasionally	Sometimes	Frequently	Usually	Always
friends							
family members							

10. **List the six people you talk to most, indicating their relationship to you. (Please use initials, no names)**

For example: co-worker J., mother, daughter, employer, friend-S., wife

Then, to the right, please rate how often you mix languages with each of these 6 people:

For instance, if your co-worker J. is #1, rate how often you mix languages with John.

		Never	Rarely	Occasionally	Sometimes	Frequently	Usually	Always
1								
2								
3								
4								
5								
6								

Did you switch between your languages within a conversation or mix your languages...

		Never	Rarely	Occasionally	Sometimes	Frequently	Usually	Always
11.	Before you turned 5 years old							
12.	During elementary and middle school							
13.	During high school							
14.	After high school until now							
15.	During the most recent five years of your life							

Please rate how often you do each of the following:

		Never	Rarely	Occasionally	Sometimes	Frequently	Usually	Always
16.	Do you use *both* Spanish and English every day?							
17.	How often do you spend the how day *without* speaking Spanish?							
18.	How often do others switch between languages within a conversation when speaking to you?							

19. **How many years of your life have you spent mixing your languages?** ______ (enter a number in years)

20. **How old are you?** ______

PLEASE RATE HOW MUCH YOU AGREE WITH THE FOLLOWING STATEMENTS.

Please rate all the of the following:

		Strongly Disagree	Moderately Disagree	Mildly Disagree	Don't Agree or Disagree	Mildly Agree	Moderately Agree	Strongly Agree
21.	I often use Spanish and English in the same conversation.							
22.	I never mix my languages.							

A code-switcher is someone who uses two languages in the same conversation.

FOR EXAMPLE: Sometimes I change only one palabra in a sentence. Other times, I start a sentence in English y termino en español. Or I might speak several sentences in English. Y luego cambio a español para acabar la conversación.

23. **Are you a code-switcher? (see above)**

Never	Rarely	Occasionally	Sometimes	Frequently	Usually	Always

Once you complete this part and begin Part 2, do not go back and change answers on Part 1.

Part 2

24. **Please rate how much you agree with the following statements about switching between and mixing languages.**

	Strongly Disagree	Moderately Disagree	Mildly Disagree	Don't Agree or Disagree	Mildly Agree	Moderately Agree	Strongly Agree
It is a way to identify with a particular culture.							
It is a way to express wo you are.							
My friends mix their languages.							
It displays a distinct multicultural identity.							
People mix languages because they enjoy it.							
I wish I had more opportunities to mix my languages.							

I switch between language because...

		Never	Rarely	Occasionally	Sometimes	Frequently	Usually	Always
25.	I enjoy it.							
26.	A a part of a cultural identity.							
27.	I don't know the word.							
28.	No similar word exists in the other language.							
29.	I can't remember the word.							
30.	Finding another suitable expression is difficult.							
31.	To convey intimacy or emotion.							
32.	It's easier or faster.							
33.	To add emphasis.							
34.	To have privacy.							

APPENDIX E:

L2 Language History Questionnaire 2.0

See http://blclab.org/ for online use and credit

Participant ID: ____________________

1. Age (in years): __________

2. Sex (Circle one): Male / Female

3. Education (your current or most recent educational level, even you have not finished the degree) (Circle one):

- Graduate school (PhD/MD/JD) • High school
- Graduate school (Masters) • Middle school
- College (BA/BS) • Other (specify):

4. Have you ever studied or learned a second language in terms of listening, speaking, reading, or writing? (Circle one):

   Yes / No

5. Indicate your native language(s) and any other languages you have studied or learned, the age at which you started using each language in terms of listening, speaking, reading, and writing, and the total number of years you have spent using each language.

Language	Listening	Speaking	Reading	Writing	Years of use[a]

a. You may have learned a language, stopped using it, and then started using it again. Please give the total number of years.

6a. Country of residence: ____________

6b. Country of origin: ____________

6c. If 6a and 6b are different, then when did you first move to the country where you currently live?

____

7. If you have lived or travelled in countries other than your country of residence or country of origin for three or more months, then indicate the name of the country, your length of stay, the language you used, and the frequency of your use of the language for each country.

Country	Length of stay[a] [month(s)]	Language	Frequency of use[b]
			1 2 3 4 5 6 7
			1 2 3 4 5 6 7
			1 2 3 4 5 6 7
			1 2 3 4 5 6 7

a. You may have been to the country on multiple occasions, each for a different length of time. Add all the trips together.

b. Please rate according to the following scale (circle the number in the table)

Never	*Rarely*	*Sometimes*	*Regularly*	*Often*	*Usually*	*Always*
1	2	3	4	5	6	7

8. Indicate the age at which you started using each of the languages you have studied or learned in the following environments.

Language	At home	With friends	At school	At work	Language software	Online games

9. Indicate the language used by your teachers for instruction at each educational level. If the instructional language switched during any educational level, then also indicate the "Switched to" language.

	Language	(Switched to)
Elementary school		
Middle school		
High school		
College/university		

10. Rate your language learning skill. In other words, how good do you feel you are at learning new languages, relative to your friends or other people you know? (circle one)

Very poor	Poor	Limited	Average	Good	Very good	excellent
1	2	3	4	5	6	7

11. Rate your current ability in terms of listening, speaking, reading, and writing in each of the languages you have studied or learned. Please rate according to the following scale (circle the number in the table):

Very poor	Poor	Limited	Functional	Good	Very good	Native-like
1	2	3	4	5	6	7

Language	Listening	Speaking	Reading	Writing
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7

12. If you have taken any standardized language proficiency tests (e.g., TOEFL), then indicate the name of the test, the language assessed, and the score you received for each.

If you do not remember the exact score, then indicate an "Approximate score" instead.

Test	Language	Score	(Approximate score)

13. Rate the strength of your foreign accent for each of the languages you have studied or learned. Please rate the strength of your accent according to the following scale (circle the number in the table):

None	Very weak	Weak	Moderate	Strong	Very strong	Extreme
1	2	3	4	5	6	7

Language	Strength of accent
	1 2 3 4 5 6 7
	1 2 3 4 5 6 7
	1 2 3 4 5 6 7
	1 2 3 4 5 6 7

14. Estimate how many hours per day you spend engaged in the following activities in each of the languages you have studied or learned.

	Language:	Language:	Language:
Watching television:	____(hrs)	____(hrs)	____(hrs)
Listening to radio:	____(hrs)	____(hrs)	____(hrs)
Reading for fun:	____(hrs)	____(hrs)	____(hrs)
Reading for school/work:	____(hrs)	____(hrs)	____(hrs)
Writing emails to friends:	____(hrs)	____(hrs)	____(hrs)
Writing for school/work:	____(hrs)	____(hrs)	____(hrs)

15. Estimate how many hours per day you spend speaking with the following groups of people in each of the languages you have studied or learned.

	Language:	Language:	Language:
Family members:	____(hrs)	____(hrs)	____(hrs)
Friends[a]:	____(hrs)	____(hrs)	____(hrs)
Classmates:	____(hrs)	____(hrs)	____(hrs)
Coworkers[b]:	____(hrs)	____(hrs)	____(hrs)

a. Include significant others in this category if you did not include them as family members (e.g., married partners).

b. Include anyone in the work environment in this category (e.g., if you are a teacher, include students as co-workers).

16a. Do you mix words or sentences from different languages when you speak? *(This includes, for example, starting a sentence in one language but using a word or phrase from another language in the middle of the sentence.)* (Circle one)

Yes / No

16b. If you answered "Yes" to 16a, then indicate the languages that you mix and estimate the frequency of mixing in normal conversation with the following groups of people. Please estimate the frequency of mixing according to the following scale (circle the number in the table):

	Language 1	Language 2	Frequency of mixing
Family members			1 2 3 4 5 6 7
Friends			1 2 3 4 5 6 7
Classmates			1 2 3 4 5 6 7
Coworkers			1 2 3 4 5 6 7

17. In which language do you communicate best or feel most comfortable in terms of listening, speaking, reading, and writing in each of the following environments?

	Listening	Speaking	Reading	Writing
At home				
With friends				
At school				
At work				

18. How often do you use each of the languages you have studied or learned for the following activities? Please circle the number in the table according to the scale below.

Never	Rarely	Sometimes	Regularly	Often	Usually	Always
1	2	3	4	5	6	7

Language	Thinking	Talking to yourself	Expression emotion[a]	Dreaming	Arithmetic[b]	Remembering numbers[c]
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1234567	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1234567	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1234567	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1234567	1 2 3 4 5 6 7	1 2 3 4 5 6 7

a. This includes shouting, cursing, showing affection, etc.
b. This includes counting, calculating tips, etc.
c. This includes telephone numbers, ID numbers, etc.

19. What percentage of your friends speaks each of the languages you have studied or learned? *(The total percentage should add up to 100%.)*

Language	Percentage
	%
	%
	%
	%

20a. Do you feel that you are bicultural or multicultural? *(This includes, for example, growing up with parents or relatives from different cultures or living in different cultures for extensive periods of time.)* (Circle one)

Yes / No

20b. If you answered "Yes" to 20a, then which cultures/languages do you identify with more strongly? Rate the strength of your connection in the following categories for each culture/language. Circle the number in the table according to the following scale.

None	Very weak	Weak	Moderate	Strong	Very strong	Extreme
1	2	3	4	5	6	7

Culture/Language	Way of life	Food	Music	Art	Cities/ towns	Sports teams
	1234567	1234567	1234567	1234567	1234567	1234567
	1234567	1234567	1234567	1234567	1234567	1234567
	1234567	1234567	1234567	1234567	1234567	1234567
	1234567	1234567	1234567	1234567	1234567	1234567

21. Please comment below to indicate any additional answers to any of the questions above that you feel better describe your language background or usage.

22. Please comment below to provide any other information about your language background or usage.

APPENDIX F:

SPANISH GRAMMAR TEST

Proficiency test (45 items)

Instructions: Read the story below about two friends at a party and select the answer that best completes each sentence.

Selecciona la mejor opción entre las que se ofrecen para cada espacio en blanco.

Creo que es muy interesante **_____** de los hábitos alimenticios de la gente. Yo, por mi

(1) a. hablo
b. hablar
c. hablando

parte, **_____** vegetariana. Cuando voy a eventos sociales, como por ejemplo fiestas, bodas o

(2) a. soy
b. estoy
c. tengo

bailes, espero que _____ comida vegetariana allí. Algunas personas dicen que _____ supone un

(3) a. hay
b. haya
c. sea

(4) a. le
b. los
c. les

inconveniente proveer _____, pero yo creo que no _____ que ser así. De hecho, la comida

(5) a. lo
b. la
c. le

(6) a. tiene
b. tenga
c. tengo

vegetariana es muy fácil _____ preparar. Y cuando no _____ ofrece, puede ser _____ gran

(7) a. en
b. a
c. de

(8) a. la
b. le
c. se

(9) a. un
b. una
c. el

problema. Yo recuerdo una vez que _____ a una fiesta de cumpleaños y _____ ser todo un

(10) a. fui
b. iba
c. voy

(11) a. resultó
b. resultaba
c. resulté

desastre. La fiesta _____ en la casa de un amigo, y él había invitado a mucha gente. Me

(12) a. estaba
b. era
c. había

sorprendió porque, _____ ser un chico sin trabajo, _____ una gran variedad de comida para

(13) a. entre
b. por
c. para

(14) a. tuvo
b. tenía
c. tuviera

para los invitados. Yo creo que si me _____ tocado a mí dar la fiesta, no _____

(15) a. hubiera
b. habría
c. había

(16) a. hubiera
b. habría
c. había

dado ni la mitad de lo que _____ allí. Pero pronto me _____ cuenta que él no había

(17) a. era
b. había
c. hubiera

(18) a. doy
b. daba
c. di

preparado nada vegetariano. Yo no pongo problemas por ese tipo de cosas, pero una amiga

_________ sí __________ hace. __________ a quejarse en frente de todos, mientras el anfitrión

(19) a. mía
b. mi
c. de mí

(20) a. le
b. se
c. lo

(21) a. Empezó
b. Empezaba
c. Empezado

sólo _____ la escena con _____ boca abierta. Yo le dije a mi amiga que _____ de causar tanto

(22) a. miró
b. miraba
c. miraría

(23) a. su
b. una
c. la

(24) a. dejaba
b. deje
c. dejara

escándalo, pero no me presto atención. Por fin, el anfitrión dijo: “La próxima vez que tenga una

fiesta, _____ algo vegetariano.” Mi amiga se ________ muchísimo y se __________a poner muy

(25) a. prepararé
b. prepararía
c. preparara

(26) a. enfada
b. enfadó
c. enfadará

(27) a. empezó
b. empieza
c. empezará

nerviosa. Insultó al anfitrión y _____ dijo que _____ muy poco considerado. Yo pensé: “ojalá

(28) a. lo
b. la
c. le

(29) a. fue
b. era
c. estaba

nunca ________ traído a ________ chica aquí”. En principio, mi única intención era que ella

(30) a. habría
b. haya
c. hubiera

(31) a. este
b. esta
c. esto

lo _________ bien. Ahora veo que obviamente me ____________. Nunca debí ___________

(32) a. pasa
b. pasara
c. pase

(33) a. equivoqué
b. equivocaba
c. equivoco

(34) a. haberla
b. haber
c. habido

traído a mi amiga. Yo ya sabía que ella no _________ comportarse adecuadamente.

(35) a. podía
b. puede
c. podría

Antes de dejar la fiesta, le _______ a mi amiga: "¡En cuanto nos ________ ido,

(36) a. dijo
b. decía
c. dije

(37) a. habríamos
b. habremos
c. hayamos

cogeremos un taxi y _________ la fiesta en mi casa donde sí ___________ comida

(38) a. seguiremos
b. seguíamos
c. seguiríamos

(39) a. tendríamos
b. teníamos
c. tendremos

vegetariana!". Mi amiga _______ muy contenta y me agradeció que ___________ tan buena

(40) a. se pone
b. se puso
c. se ponía

(41) a. soy
b. era
c. fuera

amiga. Yo le dije que se ________ ya y que ____________ en toda la comida vegetariana que

(42) a. callaba
b. callaría
c. callara

(43) a. pensara
b. pensará
c. pensaría

________ a tener en mi casa. Finalmente, ella se ___________ y nos fuimos a casa.

(44) a. vamos
b. íbamos
c. fuimos

(45) a. disculpa
b. disculpó
c. disculpaba

APPENDIX G:

ELICITED IMITATION TASK (EIT) for L2 SPANISH

You are going to hear several sentences in English. After each sentence, there will be a short pause, followed by a tone sound {TONE}. Your task is to try to repeat exactly what you hear. You will be given sufficient time after the tone to repeat the sentence. Repeat as much as you can. Remember, DON'T START REPEATING THE SENTENCE UNTIL YOU HEAR THE TONE SOUND {TONE}. Now let's begin.

We drove to the park.

I’ll call her tomorrow night.

You can buy meat at the butcher shop.

My brother just bought a brand new computer.

Sometimes they take their dog for a walk in the park.

We're going to play volleyball at the gym that I told you about.

That was the last English sentence.

Now, you are going to hear a number of sentences in Spanish. Once again, after each sentence, there will be a short pause, followed by a tone sound {TONE}. Your task is to try to repeat exactly what you hear in Spanish. You will be given sufficient time after the tone to repeat the sentence. Repeat as much as you can. Remember, DON'T START REPEATING THE SENTENCE UNTIL YOU HEAR THE TONE SOUND {TONE}. Now let's begin.

1. Quiero cortarme el pelo.
2. El libro está en la mesa.
3. El carro lo tiene Pedro.
4. Él se ducha cada mañana.
5. ¿Qué dice usted que va a hacer hoy?
6. Dudo que sepa manejar muy bien.
7. Las calles de esta ciudad son muy anchas.
8. Puede que llueva mañana todo el día.
9. Las casas son muy bonitas pero caras.
10. Me gustan las películas que acaban bien.
11. El chico con el que yo salgo es español.
12. Después de cenar me fui a dormir tranquilo.
13. Quiero una casa en la que vivan mis animales.
14. A vosotros os fascinan las fiestas grandiosas.
15. Ella sólo bebe cerveza y no come nada.
16. Me gustaría que el precio de las casas bajara.
17. Cruza a la izquierda y después sigue todo derecho.
18. Ella ha terminado de pintar su apartamento.
19. Me gustaría que empezara a hacer más calor pronto.
20. El niño al que se le murió el gato está triste.
21. Una amiga mía cuida a los niños de mi vecino.
22. El gato que era negro fue perseguido por el perro.
23. Antes de poder salir él tiene que limpiar su cuarto.
24. La cantidad de personas que fuman ha disminuido.
25. Después de llegar a casa del trabajo tomé la cena.
26. El ladrón al que atrapó la policía era famoso.
27. Le pedí a un amigo que me ayudara con la tarea.
28. El examen no fue tan difícil como me habían dicho.
29. ¿Serías tan amable de darme el libro que está en la mesa?
30. Hay mucha gente que no toman nada para el desayuno.

APPENDIX H:

ELICITED IMITATION TASK (EIT) for ENGLISH

Vas a escuchar varias oraciones en español. Después de cada oración, va a haber una pausa, seguida por un tono {TONO}. Tiene que repetir exactamente lo que escuchas. Va a haber suficiente tiempo después del tono {TONO} para repetir la oración. Repite lo más que pueda. Acuérdese, NO EMPIEZES A REPETIR LA ORACIÓN HASTA QUE ESCUCHAS EL TONO {TONO}. Ahora se comienza.

Fuimos al parque.

La llamaré mañana.

¿Te gustan las zanahorias?

Mi hermano se compró un coche azul.

A veces van al cine después de las clases.

Él y sus amigos van a jugar al fútbol en esa cancha.

Esa fue la última oración en español.

Ahora vas a escuchar oraciones en inglés. De vuelta, después de cada oración, va a haber una pausa, seguida por un tono {TONO}. Va a haber suficiente tiempo después del tono {TONO} para repetir la oración. Repite lo más que pueda. Acuérdese, NO EMPIEZES A REPETIR LA ORACIÓN HASTA QUE ESCUCHAS EL TONO {TONO}. Ahora se comienza.

1. I have to buy some new dress shoes.
2. The red ball is in the toy box.
3. We both had a salad for lunch.
4. He likes to go jogging in the park.
5. Where did you say you were going today?
6. There are too many traffic lights in this town.
7. After dinner I took a long, peaceful walk.
8. Rain is predicted to fall tomorrow morning.
9. I enjoy reading stories with happy endings.
10. Their new house is very nice but too small for my taste.
11. The little girl who got lost in the forest is scared.
12. That restaurant is supposed to have very good food.
13. I want a nice big car to go on a road trip with my friends.
14. You really enjoy listening to country music, don’t you?
15. She just finished decorating the bedrooms of her house.
16. The easiest way to get there is by driving on Main Street.
17. The person I’m dating has a wonderful sense of humor.
18. She only orders meat dishes and never easts vegetables.
19. I wish the price of plane tickets would become affordable.
20. I hope this summer will not be as hot as it was last year.
21. A good friend of mine always volunteers at the senior center.
22. The white cat that you fed yesterday was the one who caught the pigeon.
23. Before we can go to the concert, he needs to buy another ticket.
24. The most fun I’ve ever had was when we went to the beach last summer.
25. The thief who got away from the policeman was very tall and thin.
26. Would you be so kind as to set the table while I finish making dinner?
27. The number of people who go to college is increasing every year.
28. I don’t know if the 11:30 train has left the station yet.
29. The interview wasn’t nearly as stressful as you told me it would be.
30. There are a lot of people who don’t exercise enough even though they should.

References

Abrahamsson, N., & Hyltenstam, K. (2009) Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*(2), 249-306.

Abutalebi, J., Brambati, S. M., Annoni, J. M., Moro, A., Cappa, S. F., & Perani, D. (2007) The neural cost of the auditory perception of language switches: An event-related functional magnetic resonance imagine study in bilinguals. *The Journal of Neuroscience*, *27*(50), 13762-13769.

Akker, E., & Cutler, A. (2003) Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition*, *6*(2), 81-96.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998) Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419-439.

Altmann, G. T. M., & Kamide, Y. (1999) Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247-264.

Appel, R., & Muysken, P. (1987) *Language contact and bilingualism.* London: Edward Arnold.

Arnfast, J. S., & Jørgensen, J. N. (2003) Code-switching as a communication, learning, and social negotiation strategy in first-year learners of Danish. *International Journal of Applied Linguistics*, *13*(1), 23-53.

Ashby, J., Rayner, K., & Clifton, C., Jr. (2005) Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology: Section A, Human Experimental Psychology*, *58*(6), 1065-1086.

Barr, D., Levy, R., Scheepers, C., & Tily H. J. (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Basnight-Brown, D. M., & Altarriba, J. (2007) Code-switching and code-mixing in bilinguals: Cognitive, developmental, and empirical approaches. In A. Ardila & E. Ramos (Eds.), *Speech and language disorders in bilinguals*, (pp. 69-89). New York: Nova Science Publishers, Inc.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Becker, C. A. (1980) Semantic context in visual word recognition: An analysis of semantic strategies. *Memory & Cognition*, *8*(6), 493-512.

Belazi, H. M., Rubin, E. J., & Toribio, J. A. (1994) Code switching and X-bar theory: The functional head constraint. *Linguistic Inquiry*, *25*(2), 221-237.

Bentahila, A. & Davies, E. E. (1983) The syntax of Arabic-French code-switching. *Lingua*, *59*(4), 301-330.

Bentahila, A., & Davies, E. E. (1992) Code-switching and language dominance. In R. J. Harris (Ed.) *Cognitive processing in bilinguals*, (pp. 443-458). Amsterdam: North-Holland.

Berk-Seligson, S. (1986) Linguistic constraints on intrasentential code-switching: A study of Spanish/Hebrew bilingualism. *Language in Society*, *15*(3), 313-348.

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2007) Shared syntactic representations in bilinguals: Evidence for the role of word-order repetition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 931-949.

Bernolet, S., Harsuiker, R. J., & Pickering, M. J. (2013) From language-specific to shared syntactic representations: The influence of second language proficiency on syntactic sharing in bilinguals. *Cognition*, *127*(3), 287-306.

Birdsong, D., & Vanhove, J. (2016) Age of second language acquisition: Critical periods and social concerns. In Nicoladis, E. & Montanari, S. (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency*, (pp. 163-182). APA & William de Grutyer, Gbmh.

Bobb, S. C., & Wodniecka, Z. (2013) Language switching in picture naming: What asymmetric switch costs (do not) tell us about inhibition in bilingual speech planning. *Journal of Cognitive Psychology*, *25*(5), 568-585.

Bock, J. K. (1986) Syntactic persistence in language production. *Cognitive Psychology, 18*(3), 355-387.

Bock, J. K. (1990) Structure in language: Creating form in talk. *American Psychologist*, *45*(11), 1221-1236.

Bock, J. K., & Griffin, Z. M. (2000) The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology*, *129*(2), 177-192.

Bourhis, R. Y. (1983) Language attitudes and self-reports of French-English language usage in Quebec. *Journal of Multilingual and Multicultural Development*, *4*(2-3), 163-179.

Bowden, H. W. (2016) Assessing second-language oral proficiency for research: The Spanish Elicited Imitation Task. *Studies in Second Language Acquisition*, *38*(4), 647-675.

Branigan, H. P., Pickering, M. J., & McLean, J. F. (2005) Priming prepositional-phrase attachment during comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 468-481.

Carpenter, P. A., & Just, M. A. (1983) What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes*, (pp. 275-307). New York: Academic Press.

Chang, F., Dell, G. S., & Bock, K. (2006) Becoming syntactic. *Psychological Review*, *113*(2), 234-272.

Charniak, E. (1983) Passing markers: A theory of contextual influence in language comprehension. *Cognitive Science*, *7*(3), 171-190.

Chen, B., Jia, Y., Wang, Z., Dunlap, S., & Shin, J. A. (2013) Is word-order similarity necessary for cross-linguistic structural priming? *Second Language Research*, *29*(4), 375-389.

Clahsen, H., & Felser, C. (2006a) Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3-42.

Clahsen, H., & Felser, C. (2006b) How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*(12), 564-570.

Clifton, C., Jr., Staub, A., & Rayner, K. (2007) Eye movements in reading words and sentences. In R. van Gompel, M. Fisher, W. Murray, & R. L. Hill (Eds.), *Eye movement research: A window on mind and brain*, (pp. 341-372). Oxford: Elsevier Ltd.

Costa, A., & Santesteban, M. (2004) Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*(4), 491-511.

Costa, A., Santesteban, M., & Ivanova, I. (2006) How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1057-1074.

Crain, S., & Steedman, M. (1985) On not being let up the garden path: The use of context by the psychological syntax processor. In D. Dowty, L. Karttunen, & A. Zwitcky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 320-358). Cambridge: Cambridge University Press.

de Bot, K. (1992) A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics*, *13*(1), 1-24.

Davies, M. (2002-) *Corpus del Español: 100 million words, 1200s-1900s.* Available online at http://www.corpusdelespanol.org.

Davies, M. (2008-) *The Corpus of Contemporary American English: 450 million words, 1990-present.* Available online at http://corpus.byu.edu/coca/.

Davies, M., & Face, T. L. (2006) Vocabulary coverage in Spanish textbooks: How representative is it? In N. Sagarra & A. J. Toribio, (Eds.), *Selected proceedings of the 9th Hispanic linguistic symposium,* (pp. 132-143). Somerville, MA: Cascadilla Proceedings Project.

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010) Cross-linguisitic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, *31*(3), 413-438.

Desmet, T., & Declercq, M. (2006) Cross-linguistic priming of syntactic hierarchical configuration information. *Journal of Memory and Language*, *54*(4), 610-632.

Deuchar, M. (2005) Congruence and Welsh-English code-switching. *Bilingualism: Language and Cognition*, *8*(3), 255-269.

Dewaele, J. M., & Li, W. (2014) Intra- and inter-individual variation in self-reported code-switching patterns of adult multilinguals. *International Journal of Multilingualism*, *11*(2), 225-246.

Di Sciullo, A. M., Muysken, P., & Singh, R. (1986) Government and code-mixing. *Journal of Linguistics*, *22*(1), 1-24.

Donley, B. (2010) *Aventuras: Primer Curso de Lengua Española, 3rd Edition*. Boston, MA: Vista Higher Learning.

Dussias, P., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013) When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*(2), 353-387.

Escobales, C. (2014) *Validation of the ACSES instrument for measuring use and awareness of bilingual code-switching* (Doctoral dissertation). Retrieved from ProQuest. (1565179)

Federmeier, K. D., & Kutas, M. (1999) A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469-495.

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002) The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133-146.

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003) The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, *24*(3), 453-489.

Folk, J. R., & Morris, R., K. (2003) Effects of syntactic category assignment on lexical ambiguity resolution in reading: An eye movement analysis. *Memory & Cognition*, *31*(1), 87-99.

Frazier, L., & Clifton, C., Jr. (2001) Parsing coordinates and ellipsis: Copy α. *Syntax*, *4*(1), 1-22.

Frazier, L, & Fodor, J. D. (1978) The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291-325.

Frazier, L., & Rayner, K. (1982) Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178-210.

Gardner-Chloros, P., & Edwards, M. (2004) Assumptions behind grammatical approaches to code-switching: When the blueprint is a red herring. *Transactions of the Philological Society*, *102*(1), 103-129.

Genesee, F., Nicoladis, E., & Paradis, J. (1995) Language differentiation in early bilingual development. *Journal of Child Language*, *22*(3), 611-631.

Gibson, E., & Pearlmutter, N. J. (2000) Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research*, *29*(2), 231-240.

Goodman, G. O., McClelland, J. L., & Gibbs, R. W. (1981) The role of syntactic context in word recognition. *Memory & Cognition*, *9*(6), 580-586.

Green, D. W. (1998) Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67-81.

Gringràs, R. (1974) Problems in the description of Spanish/English instrasentential codeswitching. In G. A. Bills (Ed.), *South-west areal linguistics,* (pp. 167-174). San Diego, CA: Institute for Cultural Pluralism.

Grosjean, F. (1982) *Life with two languages*. Cambridge, MA: Harvard University Press.

Grosjean, F. (2001) The bilingual's language modes. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing*, (pp. 1-22). Oxford: Blackwell.

Grüter, T., Rohde, H., & Schafer, A. J. (2014) The role of discourse-level expectations in non-native speakers' referential choices. In W. Orman & M. J. Valleau (Eds.), *BUCLD 38: Proceedings of the 38th annual Boston University Conference on Language Development*, (pp. 179-191). Somerville, MA: Cascadilla Press.

Gumperz, J. J. (1977) The sociolinguistic significance of conversational code-switching. *RELC Journal*, *8*(2), 1-34.

Hale, J. (2001) A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*, (pp. 1-8).

Hartsuiker, R. J., Beerts, S., Loncke, M., Desmet, T., & Bernolet, S. (2016) Cross-linguistic structural priming in multilinguals: Further evidence for shared syntax. *Journal of Memory and Language*, *90*, 14-30.

Hartsuiker, R. J., & Pickering, M. J. (2008) Language integration in bilingual sentence production. *Acta Psychologica*, *128*(3), 479-489.

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004) Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science*, *15*(6), 409-414.

Heredia, R. R, & Altarriba, J. (2001) Bilingual language mixing: Why do bilinguals code-switch? *Current Directions in Psychological Science*, *10*(5), 164-168.

Hopp, H. (2017) Cross-linguistic lexical and syntactic co-activation in L2 sentence processing. *Linguistic Approaches to Bilingualism*, *7*(1), 96-130.

Howes, D. H., & Solomon, R. L. (1951) Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, *41*(6), 401-410.

Inhoff, A. W., & Rayner, K. (1986) Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*(6), 431-439.

Izura, C., Cuetos, F., & Brysbaert, M. (2014) Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica*, *35*(1), 49-66.

Jackendoff, R. (2002) *Foundations of language: Brain, meaning, grammar, evolution.* Oxford: Oxford University Press.

Johnson, J. S., & Newport, E. L. (1989) Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60-99.

Joshi, A. K. (1985) Processing of sentences with intrasentential code switching. In D. R. Dowty, L. Kattunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives,* (pp. 190-205). Cambridge: Cambridge University Press.

Just, M. A., & Carpenter, P. A. (1980) A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329-354.

Kantola, L. & van Gompel, R. P. G. (2011) Between- and within-language priming is the same: Evidence for shared bilingual syntactic representations. *Memory & Cognition*, *39*(2), 276-290.

Kaschak, M. P., Kutta, T. J., & Schatschneider, C. (2011) Long-term cumulative structural priming persists for (at least) one week. *Memory & Cognition*, *39*(3), 381-388.

Kennedy, A. (1998) The influence of parafoveal words on foveal inspection time: Evidence for a processing tradeoff. In G. Underwood (Ed.), *Eye guidance in reading and scene perception*, (pp. 149-180). Oxford, England: Elsevier.

Kennedy, A., & Murray, W. S. (1987) The components of reading time: Eye movement patterns of good and poor readers. In J. K. O'Regan & A. Lévy-Schoen (Eds.), *Eye movements: From physiology to cognition*, (pp. 509-520). Amsterdam: North Holland.

Kennedy, A., Pynte, J., & Ducrot, S. (2002) Parafoveal-on-foveal interactions in word recognition. *The Quarterly Journal of Experimental Psychology: Section A, Human Experimental Psychology*, *55*(4), 1307-1337.

Koostra, G. J., van Hell, J. G., & Dijkstra, T. (2012) Priming of code-switches in sentences: The role of lexical repetition, cognates, and language proficiency. *Bilingualism: Language and Cognition*, *15*(4), 797-819.

Kroll, J. F., & Stewart, E. (1994) Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149-174.

Kuperberg, G. R., & Jaeger, T. F. (2016) What do we mean by prediction in language comprehension? *Language, Cognition, and Neuroscience*, *31*(1), 32-59.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011) A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the Brain: Using our Past to Generate a Future*, (pp. 190-207). Oxford: Oxford University Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016) lmerTest: Tests in linear mixed effects models. R package version 2.0-32. URL: https://CRAN.R-project.org/package=lmerTest

Ledoux, K., Traxler, M. J., & Swaab, T. Y. (2007) Syntactic priming in comprehension: Evidence from event-related potentials. *Psychological Science*, *18*(2), 135-143.

Lemhöfer, K., & Broersma, M. (2012) Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325-343.

Lenneberg, E. H. (1967) The biological foundations of language. *Hospital Practice*, *2*(12), 59-67.

Levine, G. S. (2003) Student and instructor beliefs and attitudes about target language use, first language use, and anxiety: Report of a questionnaire study. *The Modern Language Journal*, *87*(3), 343-364.

Levy, R. (2008) Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2014) Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition, 17*(3), 673-680.

Liceras, J. M., Fernández Fuertes, R., Perales, S., Pérez-Tattam, R., & Spradlin, K. T. (2008) Gender and gender agreement in bilingual native and non-native grammars: A view from child and adult functional-lexical mixings. *Lingua*, *118*(6), 827-851.

Liebscher, G., & Dailey-O'Cain, J. (2005) Learner code-switching in the content-based foreign language classroom. *The Modern Language Journal*, *89*(2), 234-247.

Loebell, H., & Bock, K. (2003) Structural priming across languages. *Linguistics*, *41*(5), 791-824.

Luka, B. J., & Barsalou, L. W. (2005) Structural facilitation: Mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *Journal of Memory and Language*, *52*(3), 436-459.

Macnamara, J., & Kushnir, S. L. (1971) Linguistic independence of bilinguals: The input switch. *Journal of Verbal Learning and Verbal Behavior*, *10*(5), 480-487.

MacSwan, J. (1999) *A Minimalist approach to intrasentential code switching.* New York, NY: Garland Publishing, Inc.

MacSwan, J. (2000) The architecture of the bilingual language faculty: Evidence from intrasentential code switching. *Bilingualism: Language and Cognition*, *3*(1), 37-54.

Mahootian, S., & Santorini, B. (1996) Code switching and the complement/adjunct distinction. *Linguistic Inquiry*, *27*(3), 464-479.

Marian, V., Spivey, M., & Hirsch, J. (2003) Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language*, *86*(1), 70-82.

McDonald, S. A., & Shillcock, R. C. (2001) Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, *44*(3), 295-323.

McRae, K., Hare, M., Elman, J. L., & Ferretti, T. (2005) A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, *33*(7), 1174-1184.

Meijer, P. J. A., Fox Tree, J. E. (2003) Building syntactic structures in speaking: A bilingual exploration. *Experimental Psychology*, *50*(3), 184-195.

Meng, M., & Bader, M. (2000) Ungrammaticality detection and garden path strength: Evidence for serial parsing. *Language and Cognitive Processes*, *15*(6), 615-666.

Merriam-Webster Online Dictionary (2000-) URL: http://www.merriam-webster.com.

Meuter, R. F. I., & Allport, A. (1999) Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, *40*(1), 25-40.

Meyer, D. E., & Schaneveldt, R. W. (1971) Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227-234.

Molinaro, N., Barber, H. A., & Carreiras, M. (2011) Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*(8), 908-930.

Monsell, S. (1991) The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition,* (pp. 148-197). Hillsdale, NJ: Erlbaum.

Moore, D. (2002) Code-switching and learning in the classroom. *International Journal of Bilingual Education and Bilingualism*, *5*(5), 279-293.

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002) Switching languages, switching *palabras* (words): An electrophysiological study of code switching. *Brain and Language*, *80*(2), 188-207.

Morris, R. K. (2006) Lexical processing and sentence context effects. In M. J. Traxler & M. A. Gernsbacker (Eds.), *Handbook of Psycholinguistic Research: 2nd Edition,* (pp. 377-401). USA: Academic Press.

Myers Scotton, C., & Ury, W. (1977) Bilingual strategies: The social functions of code-switching. *International Journal of the Sociology of Language*, *13*, 5-20.

Ortega, L., Iwashita, N., Rabie, S., & Norris, J. M. (1999) *A multi-language comparison of syntactic complexity measures and their relationship to foreign language oral proficiency* [Funded project]. Honolulu: University of Hawai'i at Mānoa.

Pfaff, C. (1979) Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language, 55*(2), 291-318.

Philipp, A. M., Gade, M., & Koch, I. (2007) Inhibitory processes in language switching: Evidence from switching language-defined response sets. *European Journal of Cognitive Psychology*, *19*(3), 395-416.

Pickering, M. J., & Branigan, H. P. (1998) The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*(4), 633-651.

Pickering, M. J., & Garrod, S. (2013) An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329-347.

Poplack, S. (1980) Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics*, *18*(7-8), 581-618.

Posner, M. I. (1980) Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3-25.

Potter, M. C., & Lombardi, L. (1998) Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, *38*(3), 265-282.

Poulisse, N., & Bongaerts, T. (1994) First language use in second language production. *Applied Linguistics*, *15*(1), 36-57.

R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Real Academia Española (2014) Diccionario de la lengua española; Edición del Tricentenario. URL: http://dle.rae.es/?w=diccionario.

Rakowsky, A. B. (1989) *A study of intra-sentential code-switching in Spanish-English bilinguals and second language learners* (Unpublished doctoral dissertation). Brown University, Providence, RI.

Ratcliff, R., Gomez, P., & McKoon, G. (2004) A diffusion model account of the lexical decision task. *Psychological Review*, *111*(1), 159-182.

Rayner, K. (1984) Visual selection in reading, picture perception, and visual search: A tutorial review. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X: Control of language processes*, (pp. 67-96). Hillsdale, NJ: Erlbaum.

Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004) The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z reader model. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 720-732.

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006) Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*(3), 241-255.

Rayner, K., & McConkie, G. W. (1976) What guides a reader's eye movements. *Vision Research*, *16*(8), 829-837.

Roelofs, A. (2003) Shared phonological encoding processes and representation of languages in bilingual speakers. *Language and Cognitive Processes*, *18*(2), 175-204.

Rothermich, K., & Kotz, S. A. (2013) Predictions in speech comprehension: fMRI evidence on the meter-semantic interface. *NeuroImage*, *70*, 89-100.

Sankoff, D., & Poplack, S. (1981) A formal grammar for code-switching. *Papers in Linguistics: International Journal of Human Communication*, *14*(1), 3-45.

Sarandi, H. (2015) Reexamining elicited imitation as a measure of implicit grammatical knowledge and beyond...? *Language Testing*, *32*(4), 485-501.

Schoonbaert, S., & Hartsuiker, R. J., & Pickering, M. J. (2007) The representation of lexical and syntactic information in bilinguals: Evidence from syntactic priming. *Journal of Memory and Language*, *56*(2), 153-171.

Setati, M., Adler, J., Reed, Y., & Bapoo, A. (2002) Incomplete journeys: Code-switching and other language practices in mathematics, science, and English language classrooms in South Africa. *Language and Education*, *16*(2), 128-149.

Shin, J. A., & Christianson, K. (2009) Syntactic processing in Korean-English bilingual production: Evidence from cross-linguistic structural priming. *Cognition*, *112*(1), 175-180.

Smith, N. J. & Levy, R. (2013) The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302-319.

Staub, A., & Clifton, C., Jr. (2006) Syntactic prediction in language comprehension: Evidence from either... or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 425-436.

Sunderman, G., & Kroll, J. F. (2006) First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition*, *28*(3), 387-422.

Talamas, A., Kroll, J. F., & Dufour, R. (1999) From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition*, *2*(1), 45-58.

Tanenhaus, M. K., & Brown-Schmidt, S. (2008) Language processing in the natural world. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1493), 1105-1122.

Terrell, T., Andrade, M., Egasse, J. & Muñoz, E. M. (2005) *Dos mundos: Comumicación* y comunidad, *6th Edition.* Columbus, OH: McGraw-Hill World Languages.

Thothathiri, M., & Snedeker, J. (2008) Give and take: Syntactic priming during spoken language comprehension. *Cognition*, *108*(1), 51-68.

Toribio, A. J. (2001a) Accessing Spanish-English code-switching competence. *International Journal of Bilingualism, 5*(4), 403-436.

Toribio, A. J. (2001b) On the emergence of code-switching competence. *Bilingualism: Language and Cognition, 4*(3), 203-231.

Ullman, M. T. (2001) The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(2), 105-122.

van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005) Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443-467.

Vigliocco, G., Vinson, D. P., Arciuli, J., & Barber, H. (2008) The role of grammatical class on word recognition. *Brain and Language*, *105*(3), 175-184.

Vitu, F., O'Regan, J. K., & Mittau, M. (1990) Optimal landing position in reading isolated words and continuous text. *Perception & Psychophysics*, *47*(6), 583-600.

Weber, K. & Indefrey, P. (2009) Syntactic priming in German-English bilinguals during sentence comprehension. *NeuroImage*, *46*(4), 1164-1172.

West, R. F, & Stanovich, K. E. (1986) Robust effects of syntactic structure on visual word processing. *Memory & Cognition*, *14*(2), 104-112.

White, L., & Genesee, F. (1996) How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research*, *12*(3), 233-265.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004) Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272-1288.

Wright, B., & Garrett, M. (1984) Lexical decision in sentences: Effects of syntactic structure. *Memory & Cognition*, *12*(1), 31-45.

Zagona, K. T. (2002) *The syntax of Spanish*. New York: Cambridge University Press.

Zentella, A. C. (1981) Tá bien, you could answer me en cualquier idioma: Puerto Rican codeswitching in bilingual classrooms. In R. P. Duran (Ed.), *Latino language and communicative behavior,* (pp. 109-132). Norwood, NJ: Ablex Publishing Corporation.