ABSTRACT

Title of Dissertation: LEXICAL COMPETITION IN NATIVE AND NONNATIVE AUDITORY WORD RECOGNITION

Alia Katherine Lancaster, Doctor of Philosophy, 2018

Dissertation directed by: Dr. Kira Gor, Second Language Acquisition

During auditory word recognition, lexical representations that match the input as the word unfolds are activated and compete for selection. The strength of a lexical competitor during this process depends on many factors, such a frequency of occurrence. These lexical characteristics affect competition within individuals who speak primarily one language, monolinguals (e.g., Marslen-Wilson, 1987). Within those who speak two or more languages, bilinguals, the same variables induce even stronger consequences (e.g., Bradlow & Pisoni, 1999).

In both speaker types, successfully managing lexical competition requires inhibiting lexical competitors according to some theories (e.g., McClelland & Elman, 1986; Norris, 1994). In bilinguals, lexical inhibition may be related to domain-general inhibition (e.g., Blumenfeld & Marian, 2011). This link is posited to underlie the bilingual advantage, which predicts that bilinguals are more efficient at managing lexical competition due to additional native and nonnative lexical competitors. This account

contrasts with the entrenchment hypothesis (Diependaele et al., 2013), which states that individuals with more entrenched lexicons (i.e., monolinguals) more efficiently manage lexical competition. Both theories anticipate that domain-general inhibitory control may be a resource to manage lexical competitors. The current study seeks to answer questions relating to how different speaker groups manage lexical competition and if other cognitive resources come into play.

Participants completed a visual-world task, which assessed the degree of competitor influence during target access when targets and competitors phonologically overlapped (e.g., *butter-bubble*) and the competitor was present. A phonological priming task investigated processing of a previously inhibited target in prime-target pairs with phonological overlap. Competitor strength was operationalized by frequency in both tasks, with higher-frequency cohort competitors predicted to be stronger lexical competitors. Participants also completed tasks measuring domain-general inhibitory control.

Lexical competition was more evident in the visual-world than in the phonological priming task, and bilinguals were generally more susceptible to frequency effects in their second language, as predicted by the entrenchment hypothesis. Higher English proficiency, a proxy for degree of lexical entrenchment, led to less competitor influence in bilinguals. Monolinguals outperformed bilinguals in domain-general inhibitory control, which did not exhibit any impact on the lexical competition process.

LEXICAL COMPETITION IN NATIVE AND NONNATIVE AUDITORY WORD RECOGNITION

by

Alia Katherine Lancaster

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Professor Kira Gor, Chair
Professor Robert De Keyser
Professor L Robert Slevc
Dr. Jared Linck
Professor Min Wang, Dean's Representative

Acknowledgements

First and foremost, I would like to thank my advisor and committee chair Kira Gor for her role throughout the entire PhD process. She is the reason that I chose UMD and the reason I continued to explore lexical processing and embrace terms like "phonolexical ambiguity." She is a great example of a scientist who works with compassion, contagious enthusiasm, and an ability to meaningfully interpret results that to me seem chaotic.

Thank you as well to the members of my dissertation committee. To Robert DeKeyser, whose comments are both detail- and theory-oriented. To Bob Slevc, who allowed me to be a part of his lab and exemplified true interdisciplinary collaboration. I am so grateful to Bob and the LMCL for embracing and helping me for a majority of my graduate school years. To Jared Linck, who never hesitated in agreeing to help with statistics or talk through complex model interactions. Finally, to Min Wang, whose thoughtful questions shed light on some tricky findings.

I cannot go any further without thanking my parents for being wonderful examples of the power of education. My father showed me the value of earning an advanced degree and applying it to one's career. My mother started her own PhD program after I began mine and finished before I did and showed me that anything is possible. Both my parents and my sister were unwavering in their continued support of my education, even when it took me halfway across the country.

Thank you to my husband Zeke who, for better or for worse, experienced this entire process with me. From helping me juggle wedding planning and final papers to encouraging me to keep working during the final (but seemingly endless) dissertation phase, he provided endless support. Thank you for at times reminding how important this is to me but at other times providing some perspective when I needed it.

Thank you to my graduate student friends without whom I would not have been able to navigate these past six years successfully. Chris Heffner, Alix Kowalski, Zoe Schleuter, Eric Pelzl, Rachel Adler, and Nick Panza - we are a great example of an interdisciplinary cohort of graduate students in all senses. From game nights to discussions of beers to yoga, you showed me how to balance our workload with grace and fun.

Thank you to the many work colleagues who were understanding and encouraging in equal measure. To all those at the Center for Advanced Study of Language, you gave me the courage to explore my current career path by showing me that there are smart, nice people doing amazing, applicable research outside (or at least somewhat nearby) the academy. Megan Masters and Martyn Clark – thank you for your countless conversations about the ups and downs of dissertation data collection and writing. To everyone, I sincerely thank you.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACTFL	American Council on the Teaching of Foreign Languages
AUC	Area Under the Curve
IC	Inhibitory Control
L1	First Language
L2	Second Language
LexTALE	Lexical Test for Advanced Learners of English
LHQ	Language History Questionnaire
MD	Maximum Deviation
MS	Millisecond
PM	Per Million
RQ	Research Question
RT	Reaction Time
SSRT	Stop Signal Reaction Time

.

Chapter 1: Introduction

During auditory word recognition, words are heard incrementally, which results in activation all word forms that match the input at a given moment in time (e.g., Marslen-Wilson, 1987). This situation creates the issue of multiple lexical candidates. A listener's task involves, among other processes, winnowing down the lexical candidates as more input is received and managing all the competitors that are activated during this process to select the intended word. However, not all competitors are equal in strength, and lexical characteristics (e.g., frequency) interact with listener characteristics (i.e., how many languages are represented in the lexicon) to affect competitor influence. Two theories will be discussed that predict opposing interactions of lexical and listener characteristics during auditory word recognition.

Competition during spoken word recognition

Monolingual lexical competition

During spoken word recognition, there is evidence that individuals who speak one language activate multiple lexical candidates, or a group of words with similar phonological forms (e.g., Allopenna, Magnuson, & Tanenhaus, 1998). For instance, monolingual English speakers were instructed to click on the *beaker* in a four-picture display (i.e., a visual-world task). By tracking eye movements, the researchers established that participants looked more to a picture of a *beetle* than to a picture of a *carriage*, which replaced the picture of the *beetle* in some trials. The increase in proportion of looks to the phonologically related competitor *beetle* is posited to arise from activation of both words (*beetle-beaker*) as the target word is being heard millisecond by millisecond. That is, when participants have only heard /bi/, then both words are possible targets (Allopenna et al., 1998).

The strength of competition during lexical access is influenced by many factors, among which are the individual characteristics of a word form, such as frequency and the number of competitors. Although subjective frequency (i.e., the relative regularity with which an individual person encounters a word form) is hardly possible to measure, frequency across all speakers can be approximated with spoken corpora such as SUBTLEX-US and is usually used as a proxy measure (van Heuven, Mandera, Keuleers, & Brysbaert, 2014). One method of determining competitor quantity is simply counting the number of words phonologically related to the target (e.g., neighborhood density or cohort size). The combined traits of a competitor set also play a role (e.g., neighborhood frequency in Luce & Pisoni, 1998), and all of these factors can affect the ability to both quickly and accurately access a given word and the amount of influence that same word exerts as a competitor.

Phonological neighborhood density and cohort density are two metrics for quantifying the number of lexical candidates, or competitors, that arise due to word form similarity (i.e., shared activation at the phonological level). Phonological neighbors are words that differ in one phoneme substitution, deletion, or addition in any position in the word, and phonological neighborhood density is the number of words that meet this definition for any given word (Luce & Pisoni, 1998). For instance, the words *spear* and *appear* are neighbors due to the substitution of the initial phoneme in one for another. *Scuff* and *scuffle* are neighbors due to an addition. Phonological cohorts are words that overlap initially, typically measured by the number of phonemes of overlap (e.g., three-phoneme overlap), and cohort density is the number of cohort members within a word's cohort (Marslen-Wilson & Welsh, 1978). For instance, *scuffle* is one of the cohort members of the one-, two-, three-, and four-phoneme cohort of *scuff*. However, *appear* is not within any cohort of *spear* due to lack of initial overlap.

One of the first studies to systematically manipulate and examine the effects of neighborhood density with native English-speaking monolinguals was Luce and Pisoni (1998). Their article also presents the Neighborhood Activation Model (NAM), which hypothesizes neighborhood activation from bottom-up input. In a lexical decision task, words from dense neighborhoods were named more slowly than words from sparse neighborhoods, controlling for frequency, suggesting that both frequency and the number of competitors affect spoken word recognition. The same pattern occurred during a naming task, but because lexical access and competition in production and perception may be influenced by different factors (Gollan, Sandoval, & Salmon, 2011), I will only discuss perception presently.

Instead of focusing on the phonological representation of the entire word, as NAM does, the Cohort Model focuses on the competition processes as they unfold over time. It specifies that all cohort members that match the input are activated upon hearing that input at any and all points of hearing a word, so that *cat* and all other words that begin with the same two phonemes are activated upon hearing /kæ/ (Marslen-Wilson & Welsh, 1978). Cohort size impacts lexical access similarly to neighborhood density, such that words with larger cohorts, or onset density, are produced and perceived slower than words with smaller cohorts (Vitevitch, 2002; Vitevitch, Armbrüster, & Chu, 2004).

Therefore, the cost for lexical retrieval increases as the number of lexical competitors, whether neighbors or cohort members, rises.

Lexical frequency is another factor that affects the retrieval process. Higher-frequency words are often selected faster, whether in perception (Scarborough, Cortese, & Scarborough, 1977) or production (Oldfield & Wingfield, 1965). It is thought that higher-frequency words receive more activation during lexical selection, whether that be higher baseline levels or faster activation when triggered. This idea is supported by studies demonstrating that lower-frequency words receive boosts in activation from previous presentation more than higher-frequency words (e.g., Forster & Davis, 1984).

In terms of competition, higher-frequency words are stronger competitors, and stronger competitors require more inhibition under models of lexical access with lateral inhibition. For instance, participants in Segui and Grainger (1990) completed a visual unmasked priming task with orthographically overlapping primes which were either higher or lower in frequency relative to the target. During trials with lower-frequency primes, higher-frequency targets, and initial phonological overlap, target reaction time increased compared to trials without phonological overlap (i.e., inhibition). During trials with higher-frequency primes, lower-frequency targets, and initial phonological overlap, target reaction time did not differ from trials without phonological overlap. Therefore, higher-frequency targets were previously higher-frequency cohort competitors when the prime was heard, which led them to be inhibited because they were a strong competitor. This previous inhibition as a cohort competitor led to the increase in reaction time when the same word was subsequently a target. Conversely, at least in visual word recognition, lower-frequency cohort competitors were inhibited so weakly, or not at all, during prime access that there was no evidence of competitor inhibition during subsequent target access. One conclusion from this finding is that not all competitors are inhibited, but only those that have higher activation levels than the selected word at some point during the process.

While tasks such as lexical decision examine the impact of the size of the entire competitor pool or target word frequency during target lexical access, visual-world tasks are able to investigate how the effect of just one competitor changes as phonological neighborhood density and/or frequency also changes. Magnuson, Dixon, Tanenhaus, and Aslin (2007) uncovered fine-grained temporal effects of neighborhood density, cohort density, and frequency on word recognition in the visual-world paradigm. Native English-speaking monolinguals looked at visual-world displays with pictures of targets that varied orthogonally in these three dimensions. Higher-frequency targets received more fixations at all time points than lower-frequency targets, and the reverse pattern was found for cohort density (i.e., lower cohort density targets received more fixations than higher cohort density targets). Measuring proportion of looks to the target without a related competitor provides baseline evidence about the amount of word form activation during spoken word recognition. Words with higher frequency or fewer competitors received more fixations than words with lower frequency or more competitors, which suggests that words that occur more frequently or have fewer cohort competitors are more influential as competitors.

While higher activation for higher-frequency words seems intuitive and is incorporated into many models of lexical access, the number of competitors is less pervasive. As detailed in Magnuson et al. (2007), the existence of more phonological neighbors spreads out the amount of activation. That is, having more neighbors leads to less activation per individual neighbor, causing an individual target or competitor to be more active (i.e., receive more looks) if it has fewer neighbors. At the same time, methods like lexical decision tasks demonstrate that having more neighbors leads to slower recognition (e.g., Luce & Pisoni, 1998). Thus both the activation level of an individual competitor and the number of competitors influence lexical access.

To examine the effect of a single competitor and its frequency in relation to the target, Dahan and Gaskell (2007) asked native speakers of Dutch to complete a visual-world task in which cohort competitors were either higher or lower in frequency than the target. When considering time since word onset, higher-frequency competitors received more fixations during a trial than lower-frequency competitors. When considering time since the end of the disambiguation window established by gating, higher-frequency competitors still received more fixations, suggesting that frequency has an early and lasting effect during lexical access.

To explore the isolated effects of competitor neighborhood size and frequency, monolingual English speakers learned an artificial lexicon of words in Magnuson, Tanenhaus, Aslin, and Dahan (2003). Each word in the new lexicon had a cohort and rhyme competitor (i.e., words with overlapping rhymes, such as *spear* and *near*) and was presented at either a high or low frequency, thus neighbors also had low or high frequency. The study replicated previous findings with real words by demonstrating main effects of relative target and competitor frequency (i.e., lower-frequency competitors received fewer looks) and an interaction, such that the strongest competitors were those that were higher-frequency neighbors when presented with higher-frequency targets.

These results suggest that in a newly learned lexicon, higher-frequency neighbors, or more entrenched neighbors, are stronger competitors.

The studies with monolinguals establish that different methodologies are sensitive to different aspects of phonological neighborhood density and lexical frequency. Tasks such as lexical decision are sensitive to the fact that words with dense cohorts have more lexical competitors: participants are usually slower to respond to words with dense compared to sparse cohorts. At the same time, words in dense cohorts receive less activation per cohort member than words in sparse cohorts (i.e., activation is spread more thinly among dense cohorts). Eye tracking displays effects originating from to the level of lexical activation due to cohort size instead of the number of lexical competitors that are not onset-matched. Both methodologies are sensitive to the effects of frequency, namely, that higher-frequency words receive more activation, leading to faster selection and stronger competition.

Bilingual lexical competition

As bilinguals have lexical representations for both native (L1) and nonnative (L2) words, it is parsimonious to suggest that the same processes affect selection and competition that have been discussed in monolinguals. However, there are both listener and lexical features in L2 that are unique. Before those differences are explored, there is the issue of whether bilinguals do or do not activate words in both of their lexicons.

During comprehension and production, language selectivity has been an issue of debate, partially answered by the visual word paradigm. There is a vast literature on the issue of language selectivity, and only those studies related to the impact of phonological overlap and using the visual-world task with bilinguals are discussed. A visual-world task contains two or four pictures or objects, one of which is the target. One of the remaining pictures or objects is related to the target in some manner (e.g., phonological overlap) in critical trials. Participants are instructed to pick up or click on the target and eye and/or mouse movements are tracked. In visual-world tasks with eye tracking, the dependent variable is often the proportion of discrete eye saccades, or fixations, to the target and competitor. There are multiple dependent variables for visual-world tasks with mouse tracking, and the most commonly used are those that indicate the degree of influence a competitor asserted on the mouse trajectory on the way to the target (see *Methods* for more details).

In a series of studies, Spivey and Marian provided evidence that bilinguals activate their first language while completing a task in their second language. In their first study, Russian-English bilinguals who immigrated to the US around age 16 (but have since used primarily English – see further discussion below) completed a visual-world task in blocks of English and Russian. In critical trials, one of the competitors phonologically overlapped with the target between languages. For instance, in an English language trial one possible target was *marker*. The between-language phonologically related competitor was a picture of a *stamp*, the Russian translation of which is /marku/, the accusative case of /marka/. The phonological overlap between the English target *marker* and the Russian translation of the competitor /marku/ caused the participants to fixate more on these types of competitors than unrelated ones. This pattern occurred no matter if the trial was in English or Russian. That is, in English *marker* was the target with *stamp* (/marku/) as the competitor, or in Russian *marku* was the target with *marker* as the competitor. The authors conclude that the results are evidence of language non-selectivity in bilinguals (i.e., both lexicons are automatically activated during comprehension; Spivey & Marian, 1999).

In follow-up studies, Marian and Spivey (2003a, 2003b) demonstrated both between- and within-language competition for similar groups of bilinguals. In one study, stimuli were balanced for frequency and within-language competitors were added to the design. The Russian-English bilingual participants each completed a Russian and an English block. No matter the language of the block, participants exhibited within- (e.g., *speaker-spear*) and between- (e.g., *speaker-matches* /spit͡ʃki/) language activation of phonologically related competitors (Marian & Spivey, 2003a). To control for language mode (i.e., relative degree of L1 or L2 language activation at time of test), two groups of bilinguals completed the same task *either* in Russian or English. Again, both between- and within-language activation were induced as measured by fixations to phonological competitors, but the between-language activation was stronger when the task was in the L2, or English (Marian & Spivey, 2003b).

This last study, which was most balanced in terms of materials and language mode, is in line with studies conducted since then with different groups of bilinguals—within-language competition is usually observed, even in L2, but between-language activation is less robust (e.g., Mercier, Pivneva, & Titone, 2014). The Russian-English bilinguals in Marian and Spivey's experiments potentially had a shift in language dominance from Russian to English that made their between-language activation stronger than bilingual groups who have not had such a shift. There is also evidence that between-language competitors are more distracting in a visual-world task when phonetic aspects of both languages were present. Spanish-English bilinguals fixated more on English between-language competitors (e.g., a picture of *pliers*) when the Spanish target (e.g., *playa* or *beach*) was spoken with the English word-initial voice onset time than when the Spanish target was spoken with the Spanish word-initial voice onset time (Ju & Luce, 2004). Therefore, it may be that between-language competition completely affects lexical access only for situations or participants in which both languages are present, such as overlapping phonetics or a language dominance shift, and less so for situations when one language is more present or stronger than another.

Once it had been established that a bilingual's languages are both partially active during lexical access, not only from visual-world studies but also from single-word lexical decision tasks (e.g., Dijkstra, Grainger, & Van Heuven, 1999) and sentence reading measured by eye tracking (Pivneva, Mercier, & Titone, 2014), other elements, such as the proficiency of a bilingual's languages and the degree of phonological overlap, were explored. For instance, language order and proficiency impact activation such that completing visual-world trials with cognates (i.e., words that overlap in meaning and form) activated between-language competition for participants in Blumenfeld and Marian (2007) no matter the language of test, L1 or L2. However, non-cognates were activated only for participants performing the task in L2, demonstrating more between-language activation when processing a second, and possibly weaker, language. In terms of lexical characteristics, Marian, Blumenfeld, and Boukrina (2008) revealed that neighborhood density of within- and between-language competitors influenced activation, but more so for participants who completed the task in L1. For a different group of bilinguals, between-language competition was overall less strong than within-language and was modulated by age of acquisition, so much so that late bilinguals lacked any evidence of between-language competition (Canseco-Gonzalez et al., 2010). Thus, bilingual spoken word recognition is affected by phonological overlap within L1, L2, and sometimes between languages, and the visual world paradigm is effective in capturing the influence of lexical competition.

If bilinguals' lexicons are mixed at all levels such that word forms from L1 and L2 are all represented together at the lexical level, as is suggested by work such as Marian and Spivey (2003a), then neighborhood and/or cohort density are expected to increase relative to a monolingual. Namely, because bilinguals know two languages, they will have relatively more words that contain the same initial one, two, or three phonemes than a monolingual, assuming these phonemes are included in the phonological inventories of both languages. Moreover, frequency behaves slightly differently in L2 than in L1, likely due to lower subjective frequency of the L2 word forms within a bilingual's lexicon.

In one of the first studies to directly investigate the effects of neighborhood density in nonnative language processing, Bradlow and Pisoni (1999) examined the effects of talker, listener, and item characteristics on native and nonnative listeners. Basing their predictions on work with monolinguals (e.g., Luce & Pisoni, 1998), they anticipated that words from sparser neighborhoods would be easier to recognize due to a smaller number of competitors. Indeed, this was what they found when native and nonnative speakers of English transcribed words that were classified as hard (lower frequency, higher neighborhood density) or easy (higher frequency, lower neighborhood density). Specifically, nonnative speakers demonstrated lower accuracy on hard words than native speakers. A similar finding has been demonstrated with a speech perception in noise task (Takayanagi, Dirks, & Moshfegh, 2002). The effect of neighborhood density on L2 perception was isolated by Imai, Walley, and Flege (2005), who established that nonnative speakers were more accurate when identifying words with smaller neighborhoods, when controlling for frequency. Using a different methodology, Schmidtke (2014) examined pupil dilation as a measure of cognitive effort in bilinguals with visual-world targets that varied in neighborhood density. Higher neighborhood density generally contributed to more effortful recognition. Thus, bilinguals, like monolinguals, recognize words with fewer competitors more accurately and/or faster in tasks that assess recognition of words in isolation (e.g., lexical decision tasks).

Marian, Blumenfeld, and Boukrina (2008) demonstrated that eye tracking is sensitive to the level of activation of a competitor in bilinguals as well as in monolinguals. Target displays included a between-language competitor (e.g., *roof*) whose German translation (e.g., *dach*) contained phonological overlap with the target (e.g., *dove*). Bilinguals completing the task in L1 fixated more on the between-language competitors with low neighborhood density than unrelated competitors, but bilinguals fixated equally on the between-language competitors with high neighborhood density and unrelated competitors. The authors suggest that there is a finite amount of activation, which spreads across an entire neighborhood during spoken word recognition. In the case of words with sparse neighborhoods, each neighbor receives more activation, thus L1 low-density words were stronger competitors because they had relatively higher activation than L1 high-density words. This account also applies to the findings of

Magnuson et al. (2007), who saw more looks to words in sparser neighborhoods in monolinguals.

Bilinguals are sensitive to L2 frequency, often even more so than L1 speakers, signifying that strength of a competitor originating from frequency also plays a role for this population. During visual word recognition, bilinguals demonstrate the traditional frequency effect (i.e., higher-frequency words are selected more quickly) when performing the task in their L1, and an even larger frequency effect in their L2 (Van Wijnendaele & Brysbaert, 2002). This increased L2 frequency effect has been replicated in other comprehension tasks (Diependaele, Lemhöfer, & Brysbaert, 2013; Gollan et al., 2011; Lemhöfer et al., 2008) as well as production tasks (Gollan, Montoya, Cera, & Sandoval, 2008; Gollan et al., 2011). Frequency effects were also captured in a visual world task for both languages in a group of bilinguals, with the effect size nearly doubling in L2 (Duyck, Vanderelst, Desmet, & Hartsuiker, 2008). Therefore, not only are bilinguals sensitive to frequency in L2, which is observable in visual-world tasks, but often more so than in L1. With this background on the components influencing monolingual and bilingual lexical competition, I will now discuss two theories in this area: the entrenchment hypothesis and the bilingual advantage.

Entrenchment

The entrenchment hypothesis is a theory of lexical access that specifies the effects of exposure (i.e., subjective frequency) and representational quality and encompasses both monolingual and bilingual populations (Diependaele, Lemhöfer, & Brysbaert, 2013). If one starts with the premise that a larger vocabulary leads to a larger search space because there are more lexical representations, then individuals with larger vocabularies, regardless of nativeness, should take longer to access a word form, due to more competitors. However, monolinguals with larger vocabularies are often more efficient than those with smaller vocabularies during lexical access, such as being less affected by word frequency (Kuperman & Van Dyke, 2013) or orthographic neighborhood size (Chateau & Jared, 2000). Stemming from these findings, the entrenchment hypothesis suggests that those with smaller vocabularies, such as some monolinguals or the L2 in bilinguals, have received less input. This lower subjective frequency can lead to lexical representations that are not well entrenched. In bilinguals, both frequency and the quality of the phonolexical representation affect the degree of entrenchment.

Frequency

The weaker frequency effect found in individuals with larger vocabularies originates from the lower frequency range, where smaller-vocabulary individuals do not have as much or any exposure to these words; corpus studies demonstrate that corpora overestimate lower-frequency ratings for everyone but more so for those with smaller vocabularies (Kuperman & Van Dyke, 2013). It is thought that with more exposure, lexical representations are strengthened, or entrenched, to such an extent that all words are accessed faster. Accordingly, different activation levels due to frequency are less impactful during the lexical access process (i.e., all words are more active), although still present to some degree. That is, words are accessed without strong influence of individual attributes, possibly because activation levels are already high.

With regards to L2, bilingual frequency effects are similar to those of monolinguals with smaller vocabularies—overall slower production or perception and a larger frequency effect in L2, even when comparing L1 and L2 within the same bilingual participants (Duyck et al., 2008; Gollan et al., 2008; Lemhöfer et al., 2008). Bilinguals tend to have smaller vocabularies in one language compared to monolinguals, both as children (Bialystok, Luk, Peets, & Yang, 2010) and adults (Bialystok & Luk, 2012), but adult bilinguals are still within the normal range of vocabulary size in their dominant language (Portocarrero, Burright, & Donovick, 2007). Bilingual vocabulary size studies, however, often only consider one of the bilinguals' two languages, thus it is possible that bilinguals may have, in effect, a larger total vocabulary size when considering both languages, with less entrenchment in the L2 lexicon.

Both vocabulary size and bilingual status were examined in Diependaele, Lemhöfer, and Brysbaert (2013), who compared monolingual and bilingual performance on a visual progressive demasking task. Vocabulary size was assessed with the Lexical Test for Advanced Learners of English (LexTALE), which is a short, non-speeded lexical decision test that was originally meant to measure vocabulary size but has also been validated as a measure of English proficiency (Lemhöfer & Broersma, 2012). L1 and L2 effects on reaction time were not present after vocabulary size was considered. Importantly, the magnitude of vocabulary size effect was the same in both populations.

A visual lexical decision task in English with English monolinguals and Dutch-English bilinguals provided similar evidence (Brysbaert, Lagrou, & Stevens, 2017). The mid-range vocabulary (i.e., approximately one through 10 instances per million in the SUBTLEX-UK corpus), but not higher or lower, demonstrated a frequency effect, namely a slower reaction time for less frequent words. The frequency effect was larger for those with smaller vocabularies, and monolinguals and bilinguals displayed no differences once vocabulary size had been accounted for, as in Diependaele et al. (2013). Thus, the authors state that the lexical entrenchment hypothesis makes two predictions:

> "(1) participants with a small vocabulary size will show a stronger frequency effect than participants with a large vocabulary size, and (2) once vocabulary size is taken into account, no more differences in frequency effect is expected between L1 and L2 speakers." (Brysbaert, Lagrou, & Stevens, 2017, p. 535)

The general idea within the lexical entrenchment hypothesis is that the mechanisms undergirding lexical processing and representation are similar, regardless of speaker status, but are affected by exposure. That is, with increased exposure comes better entrenchment, which can be operationalized as more active representations (e.g., smaller frequency effects) and higher fidelity in representation quality (see next section). Bilinguals have less exposure to L2 than L1, which leads to L2 lexical representations that are less entrenched than L1 representations. Larger L2 vocabulary is usually associated with higher L2 proficiency (Lemhöfer & Broersma, 2012), meaning that proficiency can approximate the level of L2 exposure. Specifically, bilinguals with higher proficiency will have more exposure to L2 lexical items, which leads to lexical representations that are more entrenched.

Representation quality

Considering only the size of the L2 vocabulary during lexical competition, bilinguals should have fewer lexical competitors than monolinguals, since individuals with smaller vocabularies do not know the lower-frequency competitors in a cohort or neighborhood. However, phonemic and phonetic discrimination in L2 is often not as accurate as in L1 (Best, 1995; Flege, 1995), leading to additional lexical competitors in L2 from inexact input-to-representation matches. Yet there is evidence that all competitors, especially those in L2, are not equally influential during lexical access.

Another aspect of the entrenchment hypothesis, and one of the differences between L1 and L2 lexicons, is the quality of L2 lexical representations. This idea is also captured in the weaker links hypothesis (Gollan et al., 2008) and the lexical quality hypothesis, created for visual word recognition (Perfetti, 2007). Specifically, the combination of the phonological and lexical representations, known as the phonolexical representations (Chrabaszcz & Gor, 2014; Cook, Pandža, Lancaster, & Gor, 2016), in L2 are different in such a way as to affect aspects of word recognition (e.g., encoding, matching, competition). Often, phonolexical ambiguity stems from confusable phonemes in L1 and L2, or phonemes that are different in L1 and L2 but cause learners to incur processing costs, such as slower recognition times when these confusable phonemes form a minimal word pair. For example, allophones in L1 may be separate phonemes in L2, such as the case with English /r/ an /l/ for Japanese learners of English.

In terms of lexical competition, confusable phoneme perception leads to spurious competitor activation during lexical access (e.g., Broersma & Cutler, 2008). To determine the extent of this additional lexical activation in L2 due to confusable phonemes, Dutch-English and English monolingual speakers completed visual-world trials in English, in which some competitors overlapped phonologically. The overlap sometimes contained a confusable vowel pair for the bilinguals (e.g., the first vowels in the *panda* and *pencil*) and sometimes vowels that were not confusable (e.g., *bottle* and *beetle*). Bilinguals looked at competitors containing confusable phonemes longer than non-confusable competitors (Weber & Cutler, 2004). This additional activation of a competitor has been replicated with the consonants /r/ and /l/ in a group of Japanese-English bilinguals (Cutler, Weber, & Otake, 2006) in another visual-world task, as well as in other comprehension tasks (Broersma & Cutler, 2008; Ota, Hartsuiker, & Haywood, 2009; Sebastián-Gallés, Echeverría, & Bosch, 2005). Therefore, L2 within-language competition is increased due to spurious L2 lexical activations stemming from fuzzy or imprecise phonolexical representations, specifically those containing confusable phonemes.

The spurious L2 lexical competition induced by poor phonolexical quality is asymmetric (Cutler et al., 2006; Darcy, Daidone, & Kojima, 2013; Weber & Cutler, 2006), possibly due to dominant L1 and L2 phoneme categories that develop depending on the phoneme relationship in L1 (e.g., allophonic) and L2 (e.g., phonemic), meaning that hearing *panda* activates *pencil* but not the reverse (Weber & Cutler, 2004). Learning method may also influence spurious activations, as learners who received only auditory input were symmetrical in competitor activation and learners who received both auditory and written input were asymmetrical (Escudero, Hayes-Harb, & Mitterer, 2008). Spurious L2 activation also occurs due to embedded words and across word boundaries (Broersma, 2012; Broersma & Cutler, 2008, 2011). The asymmetrical nature of spurious activation found in some studies is an indicator that not all L1-L2 confusable pairs of words will be activated upon hearing one another, reducing the scope of spurious lexical competition in bilinguals considerably.

Even lexical representations without confusable phonemes that are phonologically similar (i.e., only one or two phonemes difference) can cause spurious lexical activation.

Cook et al. (2016) used a measure of phonolexical distance, Levenshtein Distance, with the prediction that words that were more similar phonologically (e.g., *parent-parrot*), regardless of phoneme type, would be more difficult for lower-proficiency speakers to access. This prediction was upheld. Also investigating the effects of non-confusable phonemes, Rüschemeyer, Nojack, and Limback (2008) examined neighbor competition in bilinguals during cross-modal priming and ERP tasks. Bilinguals demonstrated evidence of neighbor activation (e.g., *mouse* activated *house*) while native speakers did not, and the authors suggest that bilinguals are less efficient at inhibiting competitors during spoken word recognition. Thus, phonolexical quality has implications beyond just considering confusable phonemes, which leads to increasing competitor pools sizes for L2 lexical access.

Additional L2 competitor activation should lead to greater competition during L2 lexical access. However, it is possible that the weaker representations that become spurious competitors are not strong competitors. As demonstrated in Segui and Grainger (1990), weak (i.e., lower-frequency) competitors are not inhibited. Representations that only weakly match the input may not be greatly activated and consequently are not greatly inhibited. In fact, there is evidence from phonological priming studies that weak representations are not only not inhibited during lexical competition but are processed sublexically (Cook & Gor, 2015a; Gor & Cook, 2018).

One method for assessing lexical and sublexical process is an unmasked phonological priming task, in which participants hear a prime word followed by a target word and often make a lexicality decision for the target word (for a review of native speaker effects see Dufour, 2008; for a review of nonnative speaker findings see Gor, 2018). Researchers compare target reaction times in related trials, where the prime and the target phonologically overlap (e.g., *remain – remind*), to unrelated trials, where the prime and the target are not phonologically related (e.g., *complain – remind*). Slower target reaction times in related compared to unrelated trials are referred to as inhibition (i.e., hearing the prime inhibited target access due to the target's status as a cohort competitor of the prime) and faster reactions times in related conditions are referred to as facilitation (i.e., hearing the prime facilitated target processing due to overlapping sublexical components such as phonemes or syllables).

The probability of inhibition or facilitation occurring in a phonological priming task seems to depend heavily on the lexicality of the prime. Inhibition occurs when the prime is a word, and facilitation occurs when the prime is a nonword (Slowiaczek & Hamburger, 1992). As such, lexical properties such as relative frequency (Radeau, Morais, & Segui, 1995) and neighborhood density (Dufour & Peereman, 2003) affect the degree of inhibition but not facilitation. Conversely, sublexical properties, such as phonotactic probability (Vitevitch & Luce, 1998) affect the degree of facilitation in this task (Gor, 2018).

Cook and Gor (2015) observed facilitation in nonnative Russian speakers during a phonological priming task in prime-target pairs that were lower frequency but that were also unknown to the participants, as determined by a translation and familiarity judgement task. The authors concluded that the lower-frequency, unknown representations were weak enough to cause sublexical facilitation via the phonemes shared between target and prime. Gor and Cook (2018) isolated the effect of frequency by replicating L2 facilitation in lower-frequency words. Yet, the stimuli in the same frequency range induced semantic priming, suggesting that a word form-to-meaning association is not necessarily an indication of the lexicality of a word form representation.

One possible reason for sublexical effects for words with low entrenchment is that these types of words are not strong lexical competitors due to a low match in incoming signal and representation, which is weakly represented and so can only be weakly matched. Thus even if bilinguals do have to manage spurious activation, it may be at the sublexical rather than the lexical level (Gor, 2018).

Inhibitory control

Turning to methods of managing lexical competition, some models such as TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994) employ lateral inhibition among lexical units as a mechanism to manage non-target activation. Models such as NAM (Luce & Pisoni, 1998) and the Cohort Model (Marslen-Wilson, 1987), on the other hand, employ gradient activation, and the word with the most activation is chosen. During L2 spoken word recognition, bilinguals may activate more lexical representations from multiple sources, such as L1 (Marian & Spivey, 2003b) or spurious activation of L2 competitors with similar word forms due to confusable (Weber & Cutler, 2004) or non-confusable phonemes (Cook et al., 2016). All the postulated and/or observed extra lexical activation and ensuing inhibition has led to predictions that bilinguals may need to employ other resources, such as domain-general inhibitory control, during lexical access to manage the additional lexical competition; an idea that is tied to the overall advantage bilinguals sometimes exhibit regarding executive function abilities, or abilities that aid individuals in focusing on the goal of an activity, inhibiting irrelevant information, and making correct responses. However, without considering spurious competition, Vitevitch (2012) found that combining English and Spanish corpora resulted in very few additional phonological neighbors and concluded that we do not need to posit extra mechanisms or schemata to deal with such a small increase in conflict at the lexical level, as is suggested in the bilingual advantage literature.

This bilingual advantage refers, specifically, to the finding that bilingual children and adults perform equivalently to monolinguals in tasks measuring working memory, in the sense of holding information in mind, but outperform monolinguals in tasks measuring components of executive function, in the sense of inhibiting proponent responses and attending to task-relevant information (Bialystok, 1999, 2010; Bialystok & Luk, 2012). One aspect of executive function that bilinguals sometimes display an advantage in is inhibitory control (Bialystok, Craik, & Luk, 2008b). Some researchers have looked beyond just this simple group difference and into the exact nature of the representations that are given this advantage. For instance, bilingual children appear to be faster during congruent and incongruent trials, but not control trials, for tasks assessing inhibitory control (Bialystok, 2010). Furthermore, bilinguals are better at stimulus-stimulus inhibition, measured via a nonverbal Stroop task, than stimulus-response inhibition, measured via a Simon task (Blumenfeld & Marian, 2014). As with spoken word recognition, proficiency modulates the bilingual advantage, with higher L2 proficiency bilinguals demonstrating higher performance monitoring and marginally higher response inhibition (Singh & Mishra, 2015). In neural terms, the bilingual advantage has been presented as a more efficient use of executive function networks and areas (i.e., less activation) in bilinguals in both linguistic and non-linguistic tasks, thought to occur due to a lifetime of practice using these processes in daily communication (Abutalebi et al., 2012; Marian, Chabal, Bartolotti, Bradley, & Hernandez, 2014).

However, the bilingual advantage may not be as widespread as it appears (for evidence of a publication bias but general support for the phenomenon see de Bruin, Treccani, & Della Sala, 2015; for a critical review of the phenomenon see Paap et al., 2015). Recently there have been calls to move past group comparisons of two speaker groups (monolinguals vs. bilingual) and to instead reframe investigations into bilingual inhibitory control and executive function abilities as one of neuroplasticity across the lifespan (Baum & Titone, 2014). In their review of the bilingual advantage literature, Baum and Titone (2014) also encourage researchers to examine age of onset, proficiency, and communicative environments as factors that make groups of bilinguals unique instead of homogeneous. This difference in language history, they claim, could account for some of the differences in the bilingual advantage literature. These factors, especially language dominance and proficiency, can perhaps also account for the differences in the amount of between-language co-activation in bilingual visual-world studies.

One aspect of executive function that is particularly relevant during spoken word recognition is inhibitory control. As a word is spoken over time, multiple candidates that fit the given input are activated, such as *beaker* and *beetle* after hearing /bi/ (Allopenna et al., 1998). The correct word must be chosen so that production or perception can take place and the word integrated into the rest of the message. One mechanism for correct word selection is differential activation, whereby all candidates receive some amount of activation but only the indented word receives enough to move forward (e.g., Finkbeiner, Almeida, Janssen, & Caramazza, 2006; La Heij, 2005). Another mechanism is that non-target candidates are inhibited so that the target can be selected, such as is detailed in the Inhibitory Control Model (Green, 1998). Investigations of language switching among bilinguals led to findings also related to inhibitory control. One was that there may be an asymmetry in the cost of switching such that the native language requires more inhibition (Meuter & Allport, 1999), which is consistent with the entrenchment account. The other finding was that language switching among bilinguals has also been associated with linguistic inhibition (for a review see Declerck & Philipp, 2015).

If one focuses only on the quantity but not quality of all bilingual lexical representations (i.e., ignoring the entrenchment account), bilinguals as a group may be more efficient at lexical inhibition. The idea is that bilinguals have more practice inhibiting more lexical competitors than monolingulas, and they may recruit domain-general inhibition. In fact, some researchers have found a link between these two processes (i.e., domain-general and lexical inhibition).

Blumenfeld and Marian (2011) tested English monolinguals and English-Spanish bilinguals in a visual-world task and compared their lexical inhibitory abilities to more general inhibitory abilities. Participants performed a visual-world task in English (i.e., L1 for all) that included within-language competitors (e.g., *hammer-hamper*). Immediately following some visual-world trials were negative priming trials, which included a four-picture display, but three of the four pictures were a black asterisk while the fourth one was gray. The task was to click on the gray asterisk. The location of the gray asterisk, also known as the probe, was in the same location as the competitor, target, or an unrelated filler in the previous trial. Slower responses to a probe that replaced a competitor compared to an unrelated filler were taken to indicate continual inhibition of the competitor location. Lastly, participants completed a non-verbal Stroop, with the direction of the arrow point and arrow location creating congruent and incongruent responses.

The visual-world tasks revealed that both monolinguals and bilinguals experienced within-language competition from phonologically related competitors, but bilinguals maintained target activation longer than monolinguals. The negative priming trials revealed that when probes were in the same location as phonologically related competitors, monolinguals were slower to respond than when probes replaced unrelated fillers. Bilinguals did not exhibit this difference (i.e., the location of a previous related competitor was not still inhibited). The authors contend that the results are evidence that bilinguals return to baseline activation levels faster than monolinguals because they displayed no residual inhibition of a previous competitor during subsequent negative priming trials.

During the non-verbal Stroop, monolinguals exhibited a marginally larger Stroop effect (i.e., difference between congruent and incongruent trials), suggesting this group was slightly less efficient when inhibiting irrelevant information during the task. The researchers correlated the Stroop effect with that of competitor activation (i.e., percentage of looks to a competitor minus percentage of looks to fillers) in each participant group. Stroop effects correlated with competitor activation for bilinguals but not monolinguals, suggesting that bilinguals who were better at inhibiting the irrelevant information during the Stroop task were also better at inhibiting the within-language competitor. The researchers conclude that not only do bilinguals' lexical activation levels return to baseline along a different time course than monolinguals during spoken word recognition, but that this process, for bilinguals only, is related to domain-general inhibition (Blumenfeld & Marian, 2011).

This relationship between inhibitory control generally and during lexical access has been replicated and extended to include other factors of interest. Mercier and colleagues (2014) examined the performance of English-native and French-native bilinguals in an English-only visual-world task. In this study, both between- (e.g., *feet-girl*, /fij/) and within-language competitors (e.g., *feet-field*) were included. Participants completed a wide range of executive control tasks, such as a neutral Simon task (arrows pointing up or down), an arrow Stroop such as that used in Blumenfeld and Marian (2011), a number Stroop, and Antisaccade. A principle component analysis of these tasks revealed two components, which were used to create two composite scores. The first composite score, dubbed "cognitive inhibition," was generated from scores on the Simon and Stroop tasks. The second composite score, called "oculomotor inhibition," was generated from the pure and mixed blocks of the Antisaccade task.

All bilinguals demonstrated within-language competition in reaction times and proportion of looks to the competitor. Between-language competition occurred for all bilinguals in reaction time measures, but only for French-native bilinguals, for whom the task was in L2, in proportion of looks. The cognitive control composite score modulated within-language competition for all participants, but only between-language competition for French-native bilinguals on trials that included both between- and within-language competitors. The effect of oculomotor control was greater for these same trials and for French-native bilinguals. Overall, as seen before, between-language competition was less robust and better inhibition of within-language competition was related to better domain-general competition in bilinguals.

This main finding has been expanded to other methodologies and populations. In an fMRI setting, monomodal bilinguals activated less of the classic executive function regions than monolinguals during within-language competition trials, and only bilinguals demonstrated a correlation between executive function region activation and Simon scores, both suggesting that bilinguals are more efficient when processing lexical competition (Marian et al., 2014). Moreover, bimodal bilinguals exhibited parallel language activation during visual-world tasks, indicating that perceptual overlap is not a necessary component in lexical competition. The task was in English, and more looks to a phonologically related ASL competitor were correlated with spatial Stroop effect scores in bilinguals (Giezen, Blumenfeld, Shook, Marian, & Emmorey, 2015). Providing converging evidence of a link between domain-general and lexical competition, this study provides the first strong and consistent evidence of this relationship *between* languages for bilinguals.

Eye tracking is not the only experimental technique to reveal a relationship between inhibitory control, bilingualism, and spoken word recognition. Pupillometry, which measures pupil size as an indication of effort, has shown that bilinguals were more affected than monolinguals by the frequency and neighborhood density of a word during a visual-world task (Schmidtke, 2014). Lower inhibitory control appears to also affect the phonemic level of representation, such that lower inhibitory scores in late bilinguals led to more L2-like VOT production and L2-like VOT category boundaries in perception of

L1. The relevance of this study is that lower inhibitory control produced more overlap in the phonemic inventories of L1 and L2 (Lev-Ari & Peperkamp, 2013).

Artificial language learning tasks have revealed that adult bilinguals use inhibitory control more than monolinguals both to learn a new "language" and also to reduce phonological overlap between old and new languages (Bartolotti & Marian, 2012; Kapa & Colombo, 2014). Interestingly, when both monolinguals and bilinguals learn an artificial language vocabulary equally well, bilinguals are better able to reduce the lexical competition from phonological overlap between English and the new language, as measured by eye tracking and mouse tracking (Bartolotti & Marian, 2012).

Overall, we see that lexical competition from multiple word form activation during lexical access is observed in monolinguals and bilinguals, both within and between known languages (e.g., Allopenna et al., 1998; Spivey & Marian, 1999). Research in monolinguals suggests that characteristics such as the number of competitors and frequency affects the strength of a competitor, as demonstrated in visual-world studies such as Magnuson et al. (2003), which are particularly well-suited to capture the influence of these features on target word identification and competitor influence as the process unfolds. The usual effect of frequency for monolinguals is that higher-frequency words are selected faster and looked at more often in eye-tracking, visual-world studies, with the underlying assumption that higher-frequency words receive more activation (the exact instantiation during lexical access is theory-dependent). Bilinguals demonstrate the same effect in their L1 and with an even greater difference between higher- and lower-frequency word processing in L2 (e.g., (Duyck et al., 2008; Van Wijnendaele & Brysbaert, 2002).

Regarding how bilinguals manage lexical competition, there are two accounts that have opposing predictions. The entrenchment hypothesis (Diependaele et al., 2013) postulates that individuals with less entrenched lexicons (e.g., bilinguals or monolinguals with small vocabularies) are less efficient at lexical access, being more influenced by factors such as lexical frequency. As such, bilinguals should display larger effects of competitor frequency on lexical competition than monolinguals, who have more entrenched lexicons by virtue of more L1 exposure. This account also encompasses lexical representation quality, which is compatible with findings that L2 phonolexical representations are often not very distinct (e.g., Broersma & Cutler, 2011; Cook et al., 2016) and potentially causes spurious competitor activation. However, spurious competitors are not always strong competitors (Gor & Cook, 2018).

The bilingual advantage, on the other hand, posits that bilinguals are more efficient than monolinguals at lexical access because they have been required to inhibit both L1 and L2 competitors over a lifetime. Bilinguals usually display better domain-general inhibitory control than monolinguals, and researchers in this domain suggest increased lexical competition is one mechanism driving this advantage. A parsimonious account of lexical competition suggests that monolinguals and bilinguals use the same management processes. Accordingly, there must also be some point at which monolinguals utilize domain-general inhibitory control to manage lexical competition as well. That is, the same compensatory mechanisms are at play for bilinguals and monolinguals, but the relationship between lexical and domain-general inhibitory control is only visible for bilinguals due to additional lexical competition. One strength of the entrenchment account, and where it and the bilingual advantage agree, is that lexical access and competition are similar processes no matter how many languages one knows. The differentiating factor according to the entrenchment hypothesis is the strength of the lexical representation and according to the bilingual advantage is degree of experience with lexical competition. Therefore, both accounts predict that individuals with better domain-general inhibitory control abilities will be better able to manage lexical competition, but they diverge in the comparison of monolinguals to bilinguals during this process.

In order to tease apart these two theories, the current study compares aspects of lexical competition in monolinguals and bilinguals. A visual-world task with mouse tracking investigates competitor influence with a competitor present during target access, and a phonological priming task assesses the consequences of previous inhibition on target word access. It also clarifies the role of domain-general inhibitory control in both speaker groups in instances with stronger (i.e., higher-frequency) and weaker (i.e., lower-frequency) competitors.

Chapter 2: Research questions and predictions

The current study investigated lexical competition and differences during this process between native and nonnative speakers. Specifically, the management of competitors of different strength was compared between monolingual and bilingual participants. Some researchers predict that bilinguals may be less efficient at managing competition due to weaker lexical representations (Diependaele et al., 2013), but others predict that they may be more efficient due to increased lexical competition regulation over a lifetime (Marian et al., 2014). Furthermore, domain-general inhibitory control has been linked to managing lexical competition in bilinguals (e.g., Blumenfeld & Marian, 2011), and the current study explored its use as a compensatory mechanism for lexical competition in both speaker groups. If the bilingual advantage does originate, in part, from the lexical domain, then there should be some situations in which monolinguals recruit domain-general inhibitory control to manage lexical competition. At the same time, bilinguals and monolinguals should be equivalent in a construct unrelated to executive function, such as fine-motor control. The current study attempted to create situations (i.e., strong lexical competition) in which monolinguals with better inhibitory control performed better. The research questions (RQ) that the study was designed to answer are as follows:

RQ1. Does phonological overlap induce lexical competition?

RQ2. Do speaker groups differ in domain-general inhibitory control and fine-motor control?

RQ3. Does lexical competition vary due to frequency, inhibitory control, speaker type, or their interactions?

Competitor strength was operationalized as competitor frequency, with the assumption that higher-frequency words, competitors or targets, are more active during lexical access. While there is evidence that higher-frequency competitors are stronger competitors for monolinguals (e.g., Segui & Grainger, 1990), the same is not known in bilinguals. One the one hand, the trend of larger frequency effects in bilinguals (e.g., Van Wijnendaele & Brysbaert, 2002) suggests that higher- and lower-frequency words receive different amounts of activation (i.e., lower-frequency words are less active because they are encountered less often). Thus higher-frequency words would be stronger competitors for bilinguals due to the greater activation levels than lower-frequency words. On the other hand, bilinguals may be especially skilled in managing lexical competition because they have a greater number of competitors in their combined (i.e., L1 and L2) lexicon; as a group they are less affected by competitor strength in the form of frequency.

The current study consisted of two main tasks to assess lexical competition: a visual world with mouse tracking and a phonological priming task. In the visual-world task, the main manipulation of interest was a competitor that was higher or lower in frequency than the target. For instance, a related visual-world trial contained a target picture (e.g., *bubble*) and a competitor picture with phonological overlap (e.g., *butter*). Since the competitor *butter* was lower in frequency than the target, it was expected that *butter* would act as a weak competitor. The key comparison within the related condition was the trials with higher-frequency competitors to those with lower-frequency competitors, in which the same stimuli were present but switched roles (i.e., *butter* as the target and *bubble* as the competitor). Competitors higher in frequency than targets were expected to act as stronger competitors.

Competitors and targets were reversed across different lists, which varied across participants, and in some lists the related competitor was replaced with an unrelated competitor, such as *sailor*. In mouse tracking, a mouse trajectory that deviates more towards a competitor is an indication of a stronger competitor during lexical access, and phonological overlap has been demonstrated to induce this effect (Spivey, Grosjean, & Knoblich, 2005), with parallels to proportion of fixations in eye tracking (e.g., Allopenna et al., 1998). Using the visual-world paradigm allowed for the examination of the effects of relative competitor strength during lexical selection, as both competitor and target were possible targets until the disambiguation point. Additionally, previous studies that did demonstrate a relationship between domain-general and lexical inhibitory control used this methodology, albeit with eye-tracking, thus this portion of the current study serves as a partial replication (e.g., Blumenfeld & Marian, 2011).

Mouse tracking assessed the degree of competitor influence during target access via the effect of relative frequency of a single cohort competitor. The current study intended to incorporate a priming probe trial after each visual world trial to measure the degree of competitor inhibition after target access, as seen in Blumenfeld & Marian (2011). Piloting indicated, however, that it was not possible to induce the negative location priming previously found and including both trials types attenuated the effect of phonological overlap in the visual-world trials (see Appendix A for pilot results). Since the impact of relative frequency cannot be investigated without first inducing lexical competition from phonological overlap, the priming probe trials were removed.

The visual-world trials allowed for a targeted investigation of a single competitor over the time course of recognition, during which both the competitor and target were possible targets. The priming task, on the other hand, measured the effect of previously being an inhibited competitor on subsequent word selection (Hamburger & Slowiaczek, 1996). For instance, upon hearing the prime *remain*, participants were predicted to inhibit the cohort competitor and target *remind*, leading to increased reaction times to *remind* as the target compared to trials in which the prime was phonologically unrelated to the target (e.g., *complain-remind*). In contrast to the visual-world task, competition during priming was unconstrained. When the prime was selected, the target was yet unknown and merely one of many prime competitors. The priming task was also more semantically shallow than the visual-world task, due to the lack of object referents. That is, a lexical search could be completed among the word form representations without necessitating links to semantic information. Because access in the visual-world task was via a picture instead of a word form, the semantic information was already activated.

The main manipulation in this task was the frequency of the entire prime-target pair, operationalized as target frequency. Primes were always lower frequency than the target to prompt the priming effect (Segui & Grainger, 1990), but targets ranged from low to high frequency, albeit within a range known by L2 speakers (see the *Methods* section for details). The priming task also facilitates the examination of the sublexical processing. If lower-frequency words are not well-known for bilinguals (i.e., low L2 lexical entrenchment), then lower-frequency primes will either attenuate or reverse the usual inhibition to create facilitation of a target with the related prime (Gor & Cook, 2018). A familiarity rating task ensured participants knew all the stimuli, but lower-frequency prime-target pairs were predicted to exhibit less target inhibition in monolinguals and facilitation in bilinguals.

Two participant-level variables were examined in this study with regard to their effect on lexical competition: speaker type (monolingual, bilingual) and domain-general inhibitory control ability. Speaker type was categorical, as determined by a language history questionnaire, but additional proficiency information in bilinguals was collected and investigated. Monolinguals also completed the proficiency test (Lemhöfer & Broersma, 2012), but proficiency scores were only included in bilingual-specific analyses, as the variables speaker type and proficiency overlapped considerably by design (i.e., monolinguals typically had high proficiency scores).

Three different, reliable tasks measured inhibitory control (Ito et al., 2015), which were combined in a composite score for analysis. An additional participant-level variable, fine-motor control, was inspected via a pursuit rotor task. The objective of the task was simple – keep the mouse cursor hovered over a red circle as it moves around the outline of a circle on the computer screen. The intention of this nonlinguistic and noncognitive task was to ensure the two speaker groups were similar in some measurable way. That is, both groups were anticipated to have equivalent fine-motor skills with a computer mouse.

Predictions

Table 1 provides a summary the predicted results for each research question, excluding the second research question regarding individual differences between the groups, which this section describes in detail. The phonological priming tasks have only one outcome: reaction time (RT) to the target word. The visual-world task includes four outcomes: RT, area under the curve (AUC), maximum deviation (MD), and flips.

The MouseTracker analyzer software generates four dependent measures. Spatial attraction of a competitor in the trajectory of the mouse from the *Start* button to the target are measured by the area under the curve (AUC) and the maximum deviation (MD). Both compare the idealized trajectory (i.e., straight line from *Start* button to target) to the actual trajectory. MD is the largest perpendicular deviation between these two trajectories, and it indicates the largest degree to which the actual trajectory was affected by a competitor. AUC is the area between these two trajectories, and it indicates how the actual trajectory was affected by the competitor over the entire trial. Complexity of the trajectory is measured by flips across the *x*-axis. Reaction time from the start of a trial (clicking on the *Start* button) to clicking on a picture is recorded. Test-retest reliability in previous work of all four measures (RT, AUC, MD, and flips) was high (i.e., above .75; Freeman & Ambady, 2010).

The first research question, does phonological overlap induce lexical competition, served as a manipulation check for both the visual-world and phonological priming tasks. Based on the results of Lancaster (2017), I predicted that the competitor during related mouse-tracking trials (e.g., *butter-bubble*) would influence the mouse trajectory more than during the unrelated trials (e.g., *butter-sailor*). This prediction would result in related trials exhibiting longer RT, larger MD and AUC, and more flips than unrelated trials.

In the phonological priming task, it was expected that RT to the target in related trials (e.g., *remain-remind*) would be longer than in unrelated trials (e.g., *complain-remind*). The target in related trials was a cohort competitor of the prime, which would lead to the target being inhibited when participants hear the prime. The stimuli were chosen from the top 5,000 most frequent English words to ensure the words were known to the bilingual participants. As such, all prime-target pairs were relatively high frequency, especially for monolingual speakers. Lexical competition due to cohort competitors has been well established in both task types with monolingual speakers (e.g., Dufour, 2008; Lancaster, 2017), as well as with bilingual speakers (e.g., Blumenfeld & Marian, 2011; Cook & Gor, 2015). Thus, the likelihood of basic phonological competition was high.

Research question two, do speaker groups differ in domain-general inhibitory control and fine-motor control, concerns the relationship of domain-general inhibitory control and bilingual status, which is a topic much discussed in the bilingual advantage literature (e.g., Bialystok, 2009; Paap et al., 2015). One of the main claims of the bilingual advantage is that bilinguals have better executive function than monolinguals; therefore, my prediction was that the bilingual group would have better domain-general inhibitory control than the monolinguals. There are discussions within the executive function literature as to whether inhibitory control is a separate construct from executive function or not (Miyake & Friedman, 2012), but this issue does not change the predictions. There are other issues debated within the bilingual advantage literature, however, that would. It is possible that the bilingual advantage is more nuanced than first thought, depending on factors such as age, proficiency of bilinguals, and amount of language switching (Bialystok, 2009; Treccani & Mulatti, 2015), and there are also possible issues of publication bias and methodological differences across studies (Paap et al., 2015). Since the participants in the current study were younger monolinguals and bilinguals, it is conceivable that the cognitive advantages of bilinguals would not appear in groups at their cognitive peak (i.e., monolinguals and bilinguals would perform equally well on domain-general inhibitory control tasks).

The remaining research question, does lexical competition vary due to frequency, inhibitory control, speaker type, or their interactions, asks first about the effect of frequency on lexical competition (i.e., item-level variation), then ties in participant-level variables by investigating the effect of domain-general inhibitory control and speaker type (i.e., monolingual or bilingual). This research question and accompanying analyses were applied to each task separately. To briefly review the dependent measures, competitor influence in the visual-world task was assessed with the four mouse-tracking outcomes (i.e., RT, MD, AUC, flips). During the priming task, the priming effect (i.e., the difference between related and unrelated prime-target pairs) evaluated the impact of lexical competition on subsequent word selection.

The two previously detailed accounts, that bilinguals are more or less efficient at processing lexical competition (i.e., the bilingual advantage and entrenchment hypothesis, respectively), provide different predictions for the effects of speaker (see Table 1 for a summary). Only related trials were included in analyses regarding this research question for the visual-world outcomes. That is, during trials with phonological competition, how much of a role did the competitor play? For the priming task, both related and unrelated trials were included in the analysis, as the magnitude of the priming effect is traditionally of interest (Dufour, 2008). Discussed in detail below, I predicted that lower-frequency prime-target pairs during the priming task would demonstrate facilitation instead of inhibition in bilinguals. Since these items did not require inhibition to manage, I examined priming effect changes due to inhibitory control only in higher-frequency pairs, where I *a priori* expected lexical inhibition in both speaker groups.

Starting with lexical frequency in the visual world task, I predicted that competitors that were higher frequency than the targets were more distracting than competitors that were lower frequency than targets. Specifically, higher-frequency competitors would cause slower RT to the target and the mouse trajectories to deviate more towards the competitor (i.e., larger MD, AUC, and more flips). The basis for this prediction is that higher-frequency competitors have more activation, whether that be baseline activation levels or faster activation, than the target, making higher-frequency competitors the more likely target candidate before the disambiguation point. Monolingual Dutch speakers in Dahan and Gaskell (2007) displayed more looks to higher-frequency cohort competitors when considering a trial as starting from the word onset, as is usual in visual-world eye-tracking studies, but also when considering time since the disambiguation point (as determined by gating), suggesting that higher-frequency competitors are stronger competitors for monolinguals, as predicted.

The priming task examined the ability to access a cohort competitor that was previously activated by the prime and competed with the prime, and other activated competitors, for selection. During related trials, upon hearing the prime (e.g., *remain*), the target (e.g., *remind*) was expected to be inhibited because it was a cohort member of the prime (e.g., Hamburger & Slowiaczek, 1996). As in the visual-world task, higher-frequency cohort members were predicted to be stronger competitors, which was observable in this task by the amount of target inhibition. Higher-frequency targets were higher-frequency cohort members of the prime, meaning that they receive more inhibition than lower-frequency targets when the prime is selected. As such, I predicted that higher-frequency trials would display more inhibition, or larger RT differences between related and unrelated trials, than lower-frequency trials.

Moving to inhibitory control, the overarching expectation for the study was that monolinguals and bilinguals do not qualitatively differ in how they manage lexical competition, which is a position explicitly espoused by the entrenchment account (Diependaele et al., 2013) and one implicit in the bilingual advantage literature. If one locus of the bilingual advantage is the greater amount of lexical competition bilinguals manage both for every word selection and over a lifetime, then this link between lexical and domain-general competition should be observable in monolinguals given certain conditions (i.e., strong lexical competition).

Therefore, the prediction for visual-world trials was that individuals with better inhibitory control would display less distraction towards competitors during related trials. That is, the lexical competition induced by a cohort competitor would be more quickly resolved by individuals with better inhibitory control, leading to faster target RTs and smaller mouse trajectory deviations (i.e., MD, AUC, and flips) during related trials. In terms of the strength of a competitor (i.e., frequency) during mouse tracking, the same reasoning applies. Those with better inhibitory control were predicted to be able to better manage strong lexical competitors, leading to smaller differences between higher- and lower-frequency competitors.

As with the predictions in the visual-world task, the overall participant-level prediction for the priming task is that participants who were more efficient at managing competition would display less evidence of it during this task. Thus, participants with better inhibitory control were predicted to show smaller priming effects (i.e., smaller RT differences between related and unrelated trials). Concerning frequency, the prediction unique to this task was that bilinguals were predicted to demonstrate facilitation in lower-frequency pairs. Facilitation in a priming task in bilinguals was seen in Broersma (2012) and in lower-frequency trials in Gor and Cook (2018). Therefore, I only examined higher-frequency pairs with regard to inhibitory control in bilinguals. In monolinguals, the entire frequency range was included when investigating inhibitory control. This was another reason for separate analyses per speaker group.

It is possible that the relationship between lexical and domain-general inhibitory control exists only in bilinguals, as has previously been demonstrated (e.g., Blumenfeld & Marian, 2011), and that the analysis across groups would yield no significant effects. While this relationship was predicted from the bilingual advantage literature, it is also not incompatible with the entrenchment account. Although not discussed in the literature, it is conceivable that domain-general inhibition is a mechanism driving the more efficient lexical processing in monolinguals with larger vocabularies for the same reasons that apply to bilinguals (i.e., more lexical competition leads to more efficient lexical competition management). It is also possible that because the bilinguals in previous studies were early bilinguals, the additional exposure to L2 words (i.e., increased entrenchment) and time spent inhibiting L1 and L2 words (i.e., strengthening the link between domain-general and lexical inhibition) led to the significant correlations. The current study made all attempts to recruit bilinguals of the same proficiency (i.e., advanced) but who were late L2 learners; thus there was a possibility that this link would not be observed in monolinguals nor bilinguals due to an insufficient amount of time managing lexical competition in both groups. Due to the likelihood of speaker group differences in domain-general inhibitory control, separate analysis for each group were conducted.

Typically, bilinguals are slower overall at spoken word recognition than monolinguals (e.g., Cook, 2012), which was predicted to occur in the current study as well. The effect of speaker type on lexical competition and its subsequent interactions were where the two accounts really differ, despite the fact that they both predict that bilinguals would have more lexical competition. According to the entrenchment account, individuals with more entrenched representations, monolinguals in the current study, were predicted to display smaller competitor influence during visual world related trials because they had higher quality lexical representations. A more exact match between the auditory input and a lexical representation leads to less competitors, suggesting that those with less entrenchment have more lexical competition. The bilingual advantage, on the other hand, predicted that bilinguals were the group with smaller competitor influence during visual world related trials specifically because they had more experience managing large competitor groups.

In the extreme, there is evidence that words with very low entrenchment, such as lower-frequency L2 words, produce sublexical effects (Cook & Gor, 2015), which was explored in the priming task by comparing higher- and lower-frequency word pairs. Like native speakers, L2 speakers in Cook and Gor (2015) displayed inhibition for known L2 prime-target pairs in a similar phonological priming task. For primes that were not well-known, targets received facilitation, suggesting that nonnative speakers processed these primes at the sublexical level. Under the entrenchment account, lower-frequency L2 words were not well-known, which further specifies the predictions. Not only were bilinguals predicted to display a larger priming difference between higher- and lower-frequency pairs, but they were expected to show facilitation for lower-frequency related pairs and inhibition for higher-frequency related pairs.

The interaction of speaker type and competitor frequency also highlights differences between the two accounts. Better entrenchment (i.e., higher L2 subjective frequency and higher quality representations) leads to smaller effects of frequency (Diependaele et al., 2013) and other lexical characteristics (such as neighborhood density; Chateau & Jared, 2000) in visual-word processing, regardless of bilingual or monolingual status (Brysbaert et al., 2017; Diependaele et al., 2013). Thus, those with less entrenched lexicons were predicted to show larger activation differentials between high and low frequency words. According to this account, bilinguals were predicted to exhibit larger differences between higher- and lower-frequency competitors than monolinguals in the visual-world task, and larger priming effects in higher- compared to lower-frequency word pairs.

The bilingual advantage account predicts an opposite effect; bilinguals are more efficient lexical processors (e.g., Blumenfeld & Marian, 2011; Marian et al., 2014) due to a lifetime of practice managing lexical competition and potentially also recruiting domain-general inhibitory control. In the visual world task, this account predicts that bilinguals would have a smaller outcome (i.e., RT, MD, AUC, flips) differences between trials with higher and lower frequency competitors than monolinguals. In the priming task, this account predicts that bilinguals would display smaller inhibition differences between higher- and lower-frequency word pairs than monolinguals.

The three-way interaction of competitor frequency, speaker type, and domain-general inhibitory control builds on the predictions from the two-way interactions. As previously discussed, inhibitory control is explicitly stated as a mechanism for managing lexical competition in the bilingual advantage literature but is also compatible with the entrenchment account. That is, better inhibitory control is a boon under both accounts. Under the entrenchment account, participants that have better inhibitory control and were *monolingual* should demonstrate the smallest difference between higher- and lower-frequency competitors in the mouse-tracking task. In the priming task high-frequency trials, which were predicted *a priori* to induce more lexical competition, monolinguals with better inhibitory control should display the least amount of inhibition. According to the bilingual advantage account, participants that have better inhibitory control and were *bilingual* should demonstrate the smallest differences in both tasks.

Studies that found a link between domain-general and lexical inhibitory control often included early bilinguals (e.g., Marian et al., 2014). It is likely that early bilinguals have a more entrenched L2 lexicon than late bilinguals, due to more L2 exposure and having had more time to develop a link between the two types of inhibitory control. Examining late bilinguals provided the opportunity to demonstrate a nice counterpoint and tips the predictions in favor of the entrenchment account. However it was still possible the processes underlying the bilingual advantage are at play due to the high proficiency level.

Thus, the tasks and materials were designed to answer these research questions and empirically test these predictions. The next chapter will discuss the participants and the methodology of each linguistic task, including details on the stimuli heard and seen by participants, in addition to the implementation of the domain-general inhibitory control tasks and the pursuit rotor task.

Table 1: Predictions summary for lexical competition research questions (RQ1 and RQ3)

Variable or interaction	Visual-world outcomes (RT, AUC, MD, flips)	Phonological priming RT
Phonological overlap	Related trials will induce longer RTs, larger AUC, MD, and more flips than unrelated (i.e., competitor has more influence)	Related trials will be slower than unrelated (i.e., inhibition)
Frequency	Higher-frequency competitors will induce larger outcomes than lower-frequency on related trials (i.e., more influence from stronger competitor)	Higher-frequency targets will have larger RT differences between related and unrelated trials than lower-frequency targets (i.e., more inhibition for stronger cohort competitor)
Speaker type	**Entrenchment**: Monolinguals will have reduced outcomes (i.e., better at managing current lexical competition) **Bilingual Advantage**: Bilinguals will have reduced outcomes	**Entrenchment**: Monolinguals will have smaller differences between related and unrelated trials (i.e., better at managing previous lexical inhibition) **Bilingual advantage**: Bilinguals will have smaller differences between related and unrelated trials
Inhibitory control	Participants with better inhibitory control will have reduced outcomes on related trials	Participants with better inhibitory control will have smaller differences between related and unrelated trials for HF targets
Speaker type x frequency	**Entrenchment**: Monolinguals will have smaller differences between trials with HF and LF competitors. **Bilingual advantage**: Monolinguals will have smaller differences between trials with HF and LF competitors.	**Entrenchment:** Bilinguals will display differences between related and unrelated trials for HF targets but not LF targets. **Bilingual advantage:** Both speaker types will display differences between related and unrelated trials for HF and LF targets

Variable or interaction	Visual-world outcomes (RT, AUC, MD, flips)	Phonological priming RT
Speaker type x inhibitory control	**Entrenchment**: Monolinguals with better inhibitory control will have the smallest outcomes **Bilingual advantage**: Bilinguals with better inhibitory control will have the smallest outcomes	**Entrenchment**: Monolinguals with better inhibitory control will have the smallest differences between related and unrelated trials. **Bilingual advantage**: Bilinguals with better inhibitory control will have the smallest differences between related and unrelated trials.
Inhibitory control x frequency	Participants with better inhibitory control will have smaller differences between HF and LF competitors on related trials (i.e., better at managing lexical competition with strong competitors)	**Entrenchment:** Monolinguals with better inhibitory control will have the smallest differences between related and unrelated trials in HF trials **Bilingual advantage**: In HF trials only, bilinguals with better inhibitory control will have smaller differences between related and unrelated trials.
Speaker type x inhibitory control x frequency	**Entrenchment**: Monolinguals with better inhibitory control will have the smallest difference between HF and LF competitors **Bilingual advantage**: Bilinguals with better inhibitory control will have the smallest difference between HF and LF competitors	

Note. HF refers to higher-frequency, and LF refers to lower-frequency.

Chapter 3: Methods

Procedure

The entire two-hour study was comprised of two sessions: a 10-15 minute online prescreening session and an hour and a half in-person session. During the online prescreening session, participants completed a language history questionnaire (LHQ; see Appendix B) followed by the Lexical Test for Advanced Learners of English (aka LexTALE; Lemhöfer & Broersma, 2012), an online English proficiency test. If participants were deemed eligible according to pre-established criteria, then they were invited to sign up for the in-person session. Before completing either portion of the online session, participants filled out an online consent form.

All participants completed six tasks, with the bilingual participants completing an additional familiarity task to ensure that they were familiar with the stimuli in the visual world and priming tasks. Participants proceeded through the tasks in the following order: visual-world mouse tracking, Stop Signal (i.e., first of three domain-general inhibitory control tasks), phonological priming, pursuit rotor, Stroop (i.e., second domain-general inhibitory control task), Antisaccade (i.e., third domain-general inhibitory control task), and familiarity. There were four mouse-tracking lists and two priming lists, which lead to a total of eight list combinations with the fixed task order.

In addition to breaks programmed into tasks lasting longer than 5 minutes, there were optional breaks after the Stop Signal and pursuit rotor tasks. All tasks except for the pursuit rotor included practice trials to familiarize participants with the procedure. Most monolingual participants completed the study in a little over an hour, and most bilingual participants completed the study in an hour and twenty minutes. For full detail on the procedure, see Appendix C.

A single mouse tracking trial began with a box labeled *Start* in the bottom, center of the computer screen. Once the *Start* button was clicked, two pictures appeared in the top right and left corners of the screen, equidistant from the *Start* button. The purpose of the *Start* button is to ensure the mouse begins each trial in the same place and equidistant from all pictures. After a 500 millisecond (ms) delay to allow both pictures to be recognized, participants heard the target word via headphones. Instructions specified that participants should click on the target picture as quickly and accurately as possible and begin moving the mouse as soon as possible (i.e., do not wait until the end of the word to begin moving the mouse). If participants did not begin moving the mouse within 1000 ms of hearing the target, a message appeared informing them to begin moving the mouse sooner. Trials timed out after three seconds, and 250 ms elapsed between clicking on the target or the time-out message and the beginning of the next trial. The MouseTracker software (Freeman & Ambady, 2010) recorded *x*- and *y*-coordinates every 13-16 ms along with reaction time and accuracy.

A trial during the priming task consisted of hearing two aurally presented words. Participants were instructed to indicate via keyboard button press if each target (i.e., every other presented word) is a real word in English or not. Primes and targets were separated by an inter-stimulus interval of 350ms and participants could not respond until the target word had finished playing (as in Cook & Gor, 2015). Trials timed out after 4000 ms.

Participants

Participants were divided into two groups – native English (i.e., monolinguals) and nonnative English speakers (i.e., bilinguals). The speaker group variable can also be thought of in terms of language of test. Since the tasks were conducted in English with English stimuli, the language of test was L1 for the monolingual participants and L2 for the bilingual participants.

Group status was determined by a prescreening LHQ and the LexTALE (i.e., the English proficiency test). Monolinguals had the following qualifications: 18 or older, native speaker of English, and scored a 90% or greater on the LexTALE. Native speaker for monolinguals was defined as having been raised in an English-only environment from birth and if other languages were known, learning did not begin until after age five. Although native speakers of English may indeed occupy a range on the LexTALE, a cut-off of 90% was chosen for two reasons. The first was that all but one native English speaking participant scored 90% or above in a megastudy by the test's original creators (Lemhöfer et al., 2008), and the second was that it was taken as an indication of effort (i.e., native English speakers who score low may not try their best during the study). Additionally, if potential native English participants stated in the prescreen LHQ that they spent any amount of time using another language daily, they were not invited to participate. As much as possible, the monolingual participants included individuals who were truly monolingual.

Bilingual speakers had the following qualifications: 18 or older, native speaker of a language other than English, moved to an English-speaking environment age 12 or older, not heritage speakers, and scored an 70% or greater on the LexTALE. Analyses of the test validation data indicate that a score of 70% or above is associated with a Common European Framework of Reference for Languages (CEFR) levels B2 and above (Lemhöfer & Broersma, 2012), which corresponds to Advanced Mid on the proficiency scale used by the American Council on the Teaching of Foreign Languages (ACTFL). Overall, bilinguals in the current study were advanced and late learners of English.

Eighty-six participants completed the study, 43 monolinguals and 43 bilinguals. Three bilinguals were excluded from all analysis — one because familiarity scores were not recorded, one was a simultaneous English-Spanish bilingual not indicated in the LHQ, and one had a LexTALE score below the cut-off. Hence, the analysis of 83 participants is described in the current study. Fifty-seven participants were female (N_{male} = 29), and two participants were left handed. The bilingual average age was 23.86 years (*SD* = 4.77, range = [18, 38]), and the monolingual average age was 21.00 (*SD* = 2.97, range = [18, 34]). Bilinguals were significantly older as a group than monolinguals (t = 3.34, p = 0.001).

The age-of-arrival and proficiency standards led to bilinguals that had a few more years to education than monolinguals. Twenty-five (58%) of the bilingual participants were undergraduate students, 16 (37%) were graduate students, and two (5%) were not students of any kind. The monolingual participants were comprised of a greater percentage of undergraduate students, with 39 (91%) undergraduate students, three (7%) graduate students, and one (2%) not a student of any kind. The LexTALE score ranged from 70% to 100% (*M* = 82.46%, *SD* = 8.19%) for bilingual participants and from 90% to 100% (*M* = 96.13%, *SD* = 3.17%) for monolingual participants.

The original intention was for all bilingual participants to have the same native language, Spanish. However, after three months of recruitment efforts with minimal success, it was decided to open the study to L2 English bilinguals of all native languages, with the same proficiency and age-of-arrival restrictions. One reason for the adjustment was that many native Spanish speakers that met the proficiency requirement arrived in the United States before the age required for the study. That is, they were both highly proficient in L2 and early bilinguals. Therefore, in order to investigate a new population (i.e., late learners with high proficiency), I maintained the age criteria but widened the participant pool by including bilinguals with different native languages. Including a more heterogenous bilingual group involves some disadvantages, such as the possibility that the different language learning environmental features among the L1 groups contributed to meaningful variability in their proficiency and learning profiles, but these were deemed acceptable in order to maintain the age-of-arrival restrictions.

Bilingual participants reported the following native languages: Spanish (N = 17), Chinese (N = 9; 4 of which specified Mandarin), French (N = 3), Hindi (N = 3), Korean (N = 3), Vietnamese (N = 2), Arabic (N = 1), Kannada (N = 1), and Persian (N = 1). The average age of arrival in an English-speaking environment ranged from 12 to 29 (M = 19.21 , SD = 4.13), and the time spent in this environment ranged from five months to 17 years (M = 50.24 months, SD = 47.15 months). Twenty-seven (68%) reported knowledge of a third language (L3), and the average L3 self-rated proficiency across all skills (e.g., reading, listening) was 4.98 (SD = 1.99) on a scale of one (minimal) to 10 (native-like). The average English (i.e., L2) self-rated proficiency was 8.34 (SD = 1.12) on the same scale. In terms of the number of languages known, one bilingual participant (3%)

reported knowing six languages, one (2%) reported knowing five languages, five (13%) reported knowing four languages, 20 (50%) reported knowing three languages, and the remaining 13 (33%) reported knowing their native language and English.

Materials

Visual world

Visual-world, mouse-tracking word stimuli can be seen in their entirety in Appendix D. Pairs of words (i.e., target-competitor) that overlapped in the initial two or three phonemes were assembled from a list of the top 5,000 most frequent words from the SUBTLEX-US corpus, which was developed from spoken language in television and film (Brysbaert & New, 2009). Further criteria necessitated words that were also imageable, nouns, and longer than three phonemes to allow nonnative speakers time to show competitor influence after the disambiguation point but before the end of the word. For each pair, one word was designated as higher frequency based on the log value of frequency of occurrence plus one in the SUBTLEX-US corpus (e.g., *butter* with a log frequency of 3.018) and the other as lower frequency (e.g., *bubble* with a log frequency of 2.612).

Since the operationalization of competitor strength was frequency, a minimum log frequency difference between the higher- and lower-frequency words in the pair was required in order to ensure observable effects. That is, it was assumed that the amount of activation, and subsequent competitor inhibition, due to frequency was continuous, meaning the effect may not be observable if the frequency difference between target and competitor was too small. Dahan and Gaskell (2007) also utilized the visual-world paradigm with native Dutch speakers and detected differences in proportions of eye saccades to distractors and competitors with a minimum frequency difference of log 0.200 (approximately 15 instances per million). Thus the stimuli for the current study also attempted to meet this minimum frequency difference between target and competitor, which ranged from 0.0683 to 1.357 (*M* = 0.5557, *SD* = 0.3118).

A control was selected for each pair with a frequency in between that of the higher- and lower- frequency word (e.g., *sailor* with a log frequency of 2.801). When there was a choice (i.e., when there were more than two words with initial overlap meeting the above characteristics), all effort was taken to match a target and competitor in neighborhood size and neighborhood frequency. Table 2 summarizes the lexical characteristics in Appendix D for each word type in experimental trials.

Table 2: Visual-world stimuli average (SD) lexical characteristics by word type

Word type	Frequency pm	Frequency log	Neighbors	Neighborhood frequency - log
High frequency	75.54 (89.42)	3.4150 (0.3544)	9.97 (7.27)	50.7910 (116.7314)
Low frequency	15.26 (6.50)	2.8596 (0.1640)	9.12 (6.14)	31.2571 (54.2701)
Control	30.02 (23.30)	3.1000 (0.2535)	9.39 (8.24)	20.4042 (35.7909)

Note. pm indicates the number of instances per million in the SUBTLEX-US corpus.

The selection process led to the 33 experimental pairs plus their control items listed in Appendix D. Of the 99 words, 66 (i.e., two-thirds) had corresponding pictures in the International Picture Naming Project (Szekely et al., 2004) with normed naming agreements above 0.50 (i.e., more than half of the participants agreed on one name for a picture). The remaining 33 pictures were collected via a Google search with the requirement of being black and white line drawings, as were the pictures from the

International Picture Naming Project. If possible, multiple pictures of the same word were collected. Twenty-five native English speakers (M_{age} = 19.75, SD_{age} = 1.13) performed a naming task on the new pictures, during which participants provided only the first picture name that came to mind. When possible, pictures that had a naming agreement of 0.50 or above were retained. All pictures were 300 x 300 pixels in size.

Each target word (i.e., higher-frequency, lower-frequency, and filler targets) was recorded by a female native English speaker and scaled to 79 dB. Each item contained a 50 ms buffer at the beginning and end of the word. The average duration, in ms, of control words (e.g., *sailor*) was 825.18 (*SD* = 168.52), 731.27 (*SD* = 127.82) for higher-frequency words, and 797.12 (*SD* = 114.65) for lower-frequency words.

Due to restrictions previously mentioned, not all lexical characteristics were able to be tightly controlled in the mouse-tracking stimuli. There were five known factors that vary among the experimental pairs that were balanced among lists. The first of these factors was whether or not the phonologically overlapping higher- and lower-frequency words have a disambiguation point which contains a confusable phoneme for the Spanish-English bilinguals, and it affected six of the 33 pairs. The second factor was whether or not the higher- and lower-frequency words contained the same number of syllables, and eight of the 33 pairs did not have the same number of syllables (e.g., *hand-hammer*). In terms of type and length of phonological overlap, 16 pairs had CV overlap, 12 CC, 3 CCV, and 2 CVC, meaning that 18 pairs had two-phoneme overlap and 5 had three-phoneme overlap. Although a native Spanish speaker did provide translations and cognate checks for all stimuli, I decided to include some cognates in order to retain a reasonable number of items. The resulting compromise was that for five pairs of words, both the higher- and lower- frequency words were cognates. For the remaining 28 pairs, neither word was a cognate. Thus neither the higher- nor lower- frequency word in a cognate pair had a processing advantage. Cognate status between English and the other native languages was not known. Lastly, despite the naming task, eight pairs contained words with a corresponding picture that had naming agreement below 0.5.

Relatedness and competitor frequency of a specific pair varied between participants by list, such that each participant saw one of the pair substantiations in Table 3. Seventeen filler trials were created to maintain a 1:2 related:unrelated ratio as is common in visual world tasks (e.g., Blumenfeld & Marian, 2011). That is, in each list, there were 16 or 17 related trials and 33 unrelated trials, with half the unrelated trials being experimental (e.g., *butter-sailor*) and half being fillers. Frequency was balanced according to list and relatedness, such that half the related trials contained a high frequency target (e.g., *butter-bubble*) and half contained a low-frequency target (e.g., *bubble-butter*). Each participant completed only one list, each of which contained 50 trials, not including the practice. As a result, one-third of the trials a participant completed were related, the unrelated trials being comprised of 50% fillers and 50% unrelated experimental trials. A participant completed 16 or 17 trials with a high frequency target and the same number with a low frequency target, half of each were unrelated.

Table 3: Visual-world example pair per list

List	Target	Competitor	Competitor pair type
list A	butter	bubble	HF-related
list B	bubble	butter	LF-related
list C	butter	sailor	HF-unrelated
list D	bubble	sailor	LF-unrelated

Phonological priming

Prime-target pairs for the phonological priming task are listed in Appendix E, with all pairs overlapping in at least the initial three phonemes. The 40 pairs were again assembled from the top 5,000 most frequent words from the SUBTLEX-US corpus (Brysbaert & New, 2009) that were also longer than three phonemes since at least three phoneme overlap is necessary to induce inhibition in this paradigm in native speakers (Hamburger & Slowiaczek, 1996). Pairs contained stress on the same syllable (e.g., *remain* - *remind*) and were chosen from a range of frequencies. Neither word in a pair was a Spanish-English cognate, as determined by a Spanish-English bilingual, nor were they compound words. Within a pair, the lower-frequency word always served as the prime.

The experimental manipulation of competitor strength in this task was operationalized as target (log) frequency. That is, targets that were more frequent were predicted to be inhibited more during prime access thus taking longer to retrieve as a target. Targets ranged in frequency from 2.013 to 4.113 (*M* = 2.940, *SD* = 0.438). The average difference between the log frequency of the prime and target ranged from 0.008 to 0.491 (*M* = 0.178, *SD* = 0.135). All effort was made to match the prime and target words in log frequency, number of neighbors, and neighborhood frequency (see Table 4:).

Table 4: Phonological priming stimuli average (SD) lexical characteristics by word type

Word type	Frequency pm	Frequency log	Neighbors	Neighborhood frequency - log
Prime	33.86 (49.66)	2.9317 (0.5170)	7.95 (5.87)	23.8279 (47.8156)

Target	47.42 (62.47)	3.1307 (0.4864)	8.38 (7.13)	17.8403 (25.3082)

Note. pm indicates the number of instances per million in the SUBTLEX-US corpus.

All phonological priming stimuli were recorded by the same female native English speaker who recorded the visual-world stimuli and scaled to 79 dB. Each item contained a 50 ms buffer at the beginning and end of the word. The average duration, in ms, of primes was 799.55 (*SD* = 151.86), and 817.13 (*SD* = 144.90) for targets.

The type of phoneme overlap between the target and prime was balanced across the two lists. In terms of distribution, the 40 pairs contained 16 with CCV overlap, 14 CVC, three CCC, two CVCC, two VCC, one CCVC, one CVCV, and one VCV. These overlap types amounted to 36 pairs with three-phoneme overlap and four pairs containing four-phoneme overlap. Relatedness was the other factor that was manipulated across lists. For instance, a related prime-target pair such as *remain-remind* was presented in list A, but in list B a word from another experimental pair acted as the prime (e.g., *complain-remind*).

In addition to the experimental pairs of words, 80 fillers were created and added to each list. The same fillers were present in each list since a participant only completed one list. Twenty of the fillers were comprised of unrelated real-word pairs, and 60 were comprised of a real-word prime and a nonce target. The nonce words were downloaded from the English Lexicon project's website (Balota et al., 2007) with the specification of being four characters or longer in an attempt to match the phoneme length of the real word stimuli. All nonwords from the English lexicon project were pronounceable and were created by changing one or two letters of a real English word (e.g., *crisly*). A Spanish-English bilingual confirmed that all nonce words were not Spanish words.

Therefore, in each list a participant heard 120 pairs of words. Sixty pairs contained nonce targets and the remaining half contained real-word targets. The real-word pairs were composed of 20 related pairs and 40 unrelated pairs (20 experimental unrelated and 20 filler unrelated). Therefore, half of the trials a participant heard required a *no* response (i.e., target was not a real English word) and one-third of the *yes* responses (i.e., real word target pairs) were related.

Familiarity rating

The familiarity rating task was designed for the bilingual speakers with the purpose of eliciting information on word familiarity for experimental stimuli. Each higher-frequency, lower-frequency, and control word from the visual-world task as well as each non-filler prime and target from the phonological priming task was presented visually. For each word, participants first provided a native language translation and then rated familiarity on a scale of one to five. Keystrokes appeared on the screen as translations were typed (i.e., participants were able to view their input), and instructions indicated to not include articles or accents and to not worry about spelling too much. The familiarity scale was the same used in Cook and Gor (2015): 1 – I have never seen this word before, 2 – I don't remember what this word means, 3 – I think I know this word, 4 – I know this word, 5 – I know this word very well. The L1 of some participants was not able to be typed using the Latin alphabet, therefore those participants did not enter translations.

Each trial began with a fixation point for 500 ms, followed by four seconds to provide the translation, and finally another four seconds to provide the familiarity rating.

The word was displayed during the entire trial. Pressing the *return* key after typing in the translation moved the screen to the rating portion, during which the entire scale was displayed, and a trial ended when a participant indicated the familiarity rating. The entire task took approximately 15 minutes to complete.

If unsure about a translation, participants were instructed to type their best guess. Ten nonwords from the phonological priming task were included as attention checks. That is, it is expected that translations for the nonwords would vary, if any were even provided, and ratings would be close to, if not at, one.

Language history questionnaire

An online language history questionnaire was completed by each participant (see Appendix B for paper version). In addition to basic demographic questions (e.g., age, gender, handedness), it contained topics such as age of acquisition, self-rated proficiency, and percent daily use of each known language. The main purpose of the questionnaire was as a prescreening tool and to characterize the speaker groups.

Domain-general inhibitory control

Domain-general inhibitory control was measured with three tasks: Stroop, Stop Signal, and Antisaccade, all of which were utilized in Ito et al. (2015). These tasks have been widely used in the executive function literature and have high reliability in previous studies. During the Antisaccade task, each trial began with a central fixation point. Then another cue appeared on either side of the screen (right, left) followed by a number on either side of the screen (right, left). The task was to report the number. In the initial block, the prosaccade block, the cue always appeared on the same side of the screen as the number, training participants to look at the cue. The number in the prosaccade block was displayed for 175 ms. In the following three antisaccade blocks, the cue and target did not appear on the same side of the screen. The three antisaccade blocks became progressively more difficult by decreasing the amount of time the number was displayed from 225 ms (block 1) to 200 ms (block 2) to 175 ms (block 3). The outcome measure was the average proportion of correct responses across the antisaccade blocks. The Antisaccade task evaluated the ability to inhibit the learned pairing of cue-target location.

The Stroop task required participants to indicate the color of a word or asterisk string via key press. There were three types of trials. Neutral trials were a string of asterisks in the red, blue, or green font color. Congruent trials were color words that matched the font color (e.g., *green* printed in green font color). Incongruent trials were color words that did not match the font color (e.g., *blue* printed in green font color). The task entailed one neutral block, followed by one congruent block, followed by two incongruent blocks. Participants were given practice trials with feedback before the neutral and congruent blocks to ensure the color-key-mappings (i.e., left arrow key for red, down arrow key for blue, right arrow key for green) were well established. The outcome measure was the Stroop effect, which was the average response time on the neutral trials subtracted from the average response time on the incongruent trials. The Stroop assessed the ability to inhibit the proponent response to read the word and instead name the font color.

The Stop Signal task used in the current study was freely available for download for PC and was described in detail in Verbruggen, Logan, and Stevens (2008). The main task from the participant's point of view was to specify the shape presented after a fixation cross as a square (*z* key) or circle (*/?* key). During some trials, a stop signal was heard after the shape appeared, in the form of a 750 Hz, 75 ms beep, which indicated that no response should be given for this trial (i.e., a no-go trial). As detailed in Logan (1994), the longer the delay between the stimulus presentation and the stop signal (i.e., the stop-signal delay, or SSD), the harder it was to inhibit the response. The task was programmed to automatically track accuracy on the no-go trials and, using this information, adjusted the SSD so that participants (correctly) did not respond on 50% of the no-go trials (ideally – a range of 40-60% was acceptable). The initial SSD was 250 ms, and the subsequent no-go trial SSD increased by 50 ms if the previous no-go trial did not have a response or decreased the SSD by 50ms if it did have a response.

The task began with 35 practice trials, followed by three blocks of 64 trials each. There was a mandatory ten second break between each block, during which summary statistics (e.g., average response time) were presented. Each trial began with a fixation cross for 250ms, followed by a shape for 1,250 ms or until the participant responded. The final score for the Stop Signal task was the stop-signal reaction time (SSRT). As described in Ito et al. (2015), I calculated this score as the average SSD subtracted from the median RT on the go trials across all three blocks. A larger SSRT denoted that a participant required a smaller SSD and had overall poorer domain-general inhibitory control.

Proficiency test

The English proficiency test during the prescreening session was the Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer & Broersma, 2012). It has been validated as a test of English vocabulary size and was highly related to other measures of English proficiency, in addition to being only five minutes in duration. Diependaele, Lemhöfer, & Brysbaert (2013) used LexTALE to predict performance in a visual, progressive demasking task. While the scores for bilinguals and monolinguals were non-overlapping, the LexTALE not only had more predictive power than the L1/L2 distinction, but it also had the same effect magnitude for both populations. Both monolingual and bilingual participants completed this task. The outcome was a percentage of correct responses which was automatically adjusted for the unequal number of word and nonce items, scored online by the test website (www.lextale.com).

Pursuit rotor

The pursuit rotor task was utilized to ensure similarity between the speaker groups on a construct other than domain-general inhibitory control. This task was administered via the publicly available Psychology Experiment Building Language (PEBL) software, which comes with a suite of tasks including the pursuit rotor (Mueller & Piper, 2014; Mueller, 2012). The task was to keep the mouse hovered over a red dot as it moved steadily around the outline of a circle. When the mouse cursor was placed over the red dot, the red color became brighter to indicate correct placement. There were four trials, each lasting 15 seconds, which equated to a speed of 0.13 rotations per second.

There were two possible outcomes for this task. The first was time on target, that is, how many ms (out of 15,000) the mouse cursor was correctly placed over the red dot during each trial. The second was average deviation in pixels between the mouse and the target (i.e., the red dot) and represented error. The general purpose of the pursuit rotor task was to assess visual-spatial tracking ability, a gross motor skill, for which I expected both speaker groups to be similar.

The original pursuit rotor task involved placing a stylus on a target moving at a constant speed and demonstrated that as age increased performance decreased (Ammons, 1955). The computerized, PEBL version has revealed the same patterns and indicated that error was a more sensitive measure due to greater individual variation (Piper, 2011). A digital version has also been used as a concurrent task with language production to show that performing two tasks incurs a production cost, similar to other gross motor skills such as walking (Kemper, Schmalzried, Herman, Leedahl, & Mohankumar, 2009). The next chapter explains the general approach for all analyses, then discusses the results for each research question.

Chapter 4: Results

For analyses containing both item- and participant-level data, linear or logistic multilevel models were conducted with the most recent version of the lme4 package (Bates, Maechler, Bolker, & Walker, 2016) in R (R Core Team, 2016). Multilevel models were chosen over other analysis types, such as regression, due to numerous advantages, including the ability to include subject and item variance together, to increase power, and to incorporate variables that are not completely normally distributed (Baayen, Davidson, & Bates, 2008; Linck & Cunnings, 2015). All models were run as forced-entry models for fixed effects and included cross-classified subject and item random intercepts.

For the second research question (i.e., speaker group differences in domain-general inhibitory control and fine-motor skills), multilevel models were not appropriate as there was no item-level variation to account for — there was only one score per participant (e.g., Stroop effect). In these instances, *t*-tests and correlations were conducted to assess the relationship among the measures as well as differences between speakers (i.e., monolingual, bilingual).

In the multilevel models, I took the following approach. First, I ran a fully maximal model in terms of fixed and random effects structures including only experimental design factors. Any nuisance variables, such as list or neighborhood frequency, were not added as there was no *a priori* reason to believe they would cause as much variance as the variables manipulated in the study. The visual-world experimental design variables were condition (unrelated, related), speaker (monolingual, bilingual), target-competitor frequency difference (continuous), inhibitory control (IC) composite (continuous), and LexTALE score (continuous). The phonological priming experimental design variables were condition (unrelated, related), speaker (monolingual, bilingual), target frequency (continuous), IC composite (continuous), and LexTALE score (continuous). All continuous variables were centered, and the IC composite score was additionally transformed into a *z*-score, with higher values representing better IC abilities. LexTALE was included in models with bilinguals speakers only.

Random effects structures were fully maximal unless the model did not converge. If converge issues occurred, by-participant and then by-item correlations among random effects were suppressed until convergence was achieved. The by-item random slope variables were speaker, IC composite, and LexTALE. The by-participant random slope variables comprised target frequency and condition. Each random effect term was entered if it was also entered as a fixed effect. For instance, the models answering research question one contained condition as the only fixed effect, thus it is the only by-participant random slope specified.

Contrast coding was applied to all multilevel models and the significance of each fixed effect term was tested using a Satterthwaite approximation for denominator degrees of freedom and *F*-statistic using the afex R package (Singmann, 2018). Contrast coding allows for an ANOVA-like output, such as the main effect of condition collapsing across all other factor levels and for the sample-mean value of any continuous variables. The afex R package calls on the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2018) to provide *F*- and corresponding *p*-values for each parameter using Satterthwaite's degree of freedom method (Satterthwaite, 1946) by eliminating each fixed effect from the model one-by-one. Therefore an ANOVA table output for each model is provided which answers questions such as was there a main effect of speaker on reaction times (RT).

For pairwise comparisons of fixed effects, estimated marginal means were utilized to understand the direction and specific comparisons in the significant main effects and interactions. The R package emmeans (Lenth, 2018) was employed, which allows for a Holm correction method for multiple comparisons (Holm, 1979). For example, if the main effect of speaker was significant on visual-world RTs, the estimated marginal means for the monolinguals and bilinguals were generated to determine which speaker group was slower, across the other conditions and for the sample-mean value of the continuous variables. One advantage of contrast coding was that it allowed ANOVA-like output and interpretations, but one disadvantage was that the model estimates and corresponding significance tests compare a factor level, such as monolinguals, to the intercept (e.g., the RT grand mean). Thus to understand differences between categorical groups, planned comparisons were carried out on the estimated marginal means using a *z*-value and significance test.

Thus, for each model, an ANOVA ouptut table is provided in this section detailing the significance of each fixed effect parameter (for full random and fixed effects model outputs, see Appendix F). For each significant categorical variable main effect, the difference in estimated model means is listed, along with the *z*- and *p*-value for the difference. For each continuous variable main effect, the contrast coded model estimate, standard error, degrees of freedom, *t*-value, and *p*-value are given. This model estimate can be interpreted as a slope of the effect of the continuous variable on the outcome. For each significant interaction, estimated model means or slopes by moderator variables, *z*-values, and *p*-values, are provided.

The analysis described below included only words that were familiar to the nonnative participants. Bilingual participants rated familiarity with each experimental target, competitor, and prime, on a scale of one (I have never seen this word before) to five (I know this word very well). Ten nonce words from the priming task were included to ascertain attention during the familiarity task. Across participants, 317 (77.89%) of the nonce words received a rating of one, 54 words (13.27%) received a two rating, 21 words (5.16%) received a three rating, six words (1.47%) received a four rating, and nine words (2.21%) received a five rating. The four and five ratings did not originate from a single participant or word.

Previous work utilizing the same rating scale classified words with ratings of one and two as "unknown," ratings of three as "recognized," and ratings of four and five as "well-known" (Cook & Gor, 2015a). In order to represent lexical competition only in words fully known by bilingual participants, trials in which either the target, the competitor, or the prime was rated as one or two were excluded. While the words with ratings of one or two seem clearly unfamiliar to participants, there are advantages and disadvantages for including and excluding words with a rating of three (i.e., "recognized"). On one hand, words with a three rating may not be known to a degree to cause lexical competition, in which case they should be excluded. On the other hand, the middle point of this five-point scale may represent different levels of familiarity for different participants, and retaining these words increases the number of items per participant. For the following analysis, I decided to include words with a three rating, which represents 2.43% of the visual-world words and 3.78% of the phonological priming words. Future analyses could explore translation accuracy in all five rating categories in order to determine how words with a three rating should be grouped.

In the visual-world task, excluding words with ratings of one or two led to the elimination of 3.11% of the trials, and the unknown words were relatively evenly split between conditions (21 in related trials, 18 in unrelated trials; 21 with lower-frequency competitors, 18 with higher-frequency competitors). In the priming task, 6.26% of bilingual trials were eliminated due to an unknown prime or target, and the unknown stimuli were equally split between conditions (45 related trials, 46 unrelated trials). Some words were not rated by participants during the familiarity task, likely due to the time limit of four seconds to provide a rating per word during the task. These unrated words were included in the subsequent analysis and comprise 0.71% (9) and 1.24% (18) of trials in the visual world and priming tasks, respectively.

The analyses are presented below in the following order. First, research question (RQ) two is discussed, as it is necessary to understand the differences in domain-general inhibitory control and fine-motor skills between speaker groups for the remaining findings. Second, visual-world results are examined, with RQ1, RQ3, and exploratory analyses clearly delineated. Lastly, the priming task results follow the same progression as the visual-world analyses. For both lexical competition tasks, first models with all speakers are discussed, followed by separate models for bilinguals and monolinguals. The reason for this split was that the LexTALE, or English proficiency score, was only relevant for the bilingual participants. By design the LexTALE differed between speakers in terms of value (i.e., monolinguals had a higher LexTALE on average than bilinguals) and homogeneity (i.e., monolingual LexTALE scores incorporated a smaller range).

Therefore it was not included in models with both speaker types to avoid any collinearity issues.

Individual difference measures

Of the three inhibitory control tasks, the Antisaccade score was based on proportion of correct answers and the scores for the Stroop and Stop Signal tasks were based on reaction time. Specifically, the Antisaccade score was the averaged proportion of correct responses during the three antisaccade blocks. No data trimming procedures were implemented for this task, but an arcsine transformation was applied in improve normality, as in Friedman et al. (2008). A higher score on the Antisaccade task represented a greater ability to inhibit a learned response. The average untransformed Antisaccade score was 0.70 (*SD* = 0.13). For all individual difference measures, I assessed odd-even split-half reliability using the Spearman-Brown prophecy formula adjustment (i.e., corrected correlation of scores using only odd-numbered or even-numbered items; Brown, 1910; Spearman, 1910). Reliability in the Antisaccade task was high at 0.91.

For the Stroop task, the exclusion criterion from previous research with the same task was applied (Friedman et al., 2008; Ito et al., 2015), namely excluding incorrect trials, trials less than 200ms, and trials greater or less than 3.32 the median absolute deviation (MAD) per participant, which is a robust trimming technique for nonnormal data (Wilcox & Keselman, 2003). These procedures led to 14.19% (1208) of trials excluded. The Stroop score was calculated by subtracting the average neutral (e.g., asterisks in blue ink) from the average incongruent (e.g., *green* in blue ink) RT. Two participants did not understand the task in the incongruent blocks, as they correctly answered zero and two trials respectively. These participants' Stroop scores were therefore not included in their inhibitory control composite scores. The average Stroop effect score was 80.31 ms (*SD* = 74.39), and the odd-even reliability score of 0.92.

The Stop Signal score was known as the Stop Signal Reaction Time (SSRT), which was the difference between the untrimmed median RT of the go trials and the stop signal delay value across all blocks (Ito et al., 2015). One participant had very low accuracy, well below chance of 0.50, on the no-go trials, suggesting that they did not establish the correct mapping of response key to shape (i.e., *z*-key for square, *?*-key for circle). Therefore, I did not include the SSRT score in their inhibitory control composite. This participant was not one of the participants whose Stroop score was excluded from their inhibitory control composite. The average SSRT was 536.90 ms (*SD* = 134.62), and reliability on SSRT scores was high at 0.98. The scoring procedures for both RT tasks produced a score whereby a smaller number indicated better inhibition. Accordingly, both were reversed so that a larger number represented better inhibition and matched the directionality of Antisaccade.

Fine-motor control was measured with a pursuit rotor task, the outcomes of which were average deviation in pixels from the target and average time on target in ms over four trials. One participant did not complete this task due to a technical error, thus does not have this score included in their IC composite score. This participant was not the same as the three previously mentioned that did not have a Stroop or SSRT score included in their IC composite. No trimming procedures were applied. The average deviation across the four trials was 3,329.82 pixels (*SD* = 5155.26), and the average time on target across the four trials was 6,493.09 ms (*SD* = 4,733.33). Reliability was respectable for each outcome, 0.989 for deviation and 0.997 for time on target. Neither outcome was normally distributed, but the Shapiro-Wilks test statistic indicated a more normal distribution for deviation, 0.87 compared to 0.63 for time on task, therefore deviation was chosen as the outcome to use from this task when comparing speaker groups.

RQ2

The three inhibitory control tasks were *z*-scored and combined to create a composite inhibitory control score (IC composite) for each participant, which was then itself *z*-scored. The three tasks were significantly correlated in Ito et al. (2015), although not very strongly as each is predicted to assess a different aspect of inhibitory control. It was expected that monolinguals and bilinguals would perform equivalently on the pursuit rotor task, and bilinguals would outperform monolinguals in domain-general inhibitory control.

In the current study, the Antisaccade and Stop Signal tasks were correlated with one another, and the Stroop task did not correlate with either (see Table 5). The pursuit rotor outcome of pixel deviation correlated with all but the Stroop score.

Table 5: Correlations among individual difference measures

	1	2	3	4
1. Antisaccade (asin)	-			
2. Stop signal (reversed)	0.25*	-		
3. Stroop (reversed)	0.15	0.03	-	
4. Pixel deviation	-0.40**	-0.26*	-0.10	-
5. IC composite	0.73**	0.66**	0.56**	-0.39**

Notes. ***p < 0.01, **p <.01, *p <.05, +p <.1

T-tests were conducted to examine differences between the two speaker groups in terms of IC composite and deviation scores. Monolinguals displayed a significantly higher IC composite than bilinguals ($t = 3.63$, $p < 0.001$), suggesting they had better domain-general inhibitory control abilities. Monolinguals also had a significantly lower deviation score ($t = -2.88$, $p = 0.005$), meaning that their fine-motor skills were better in this task (see Figure 1).

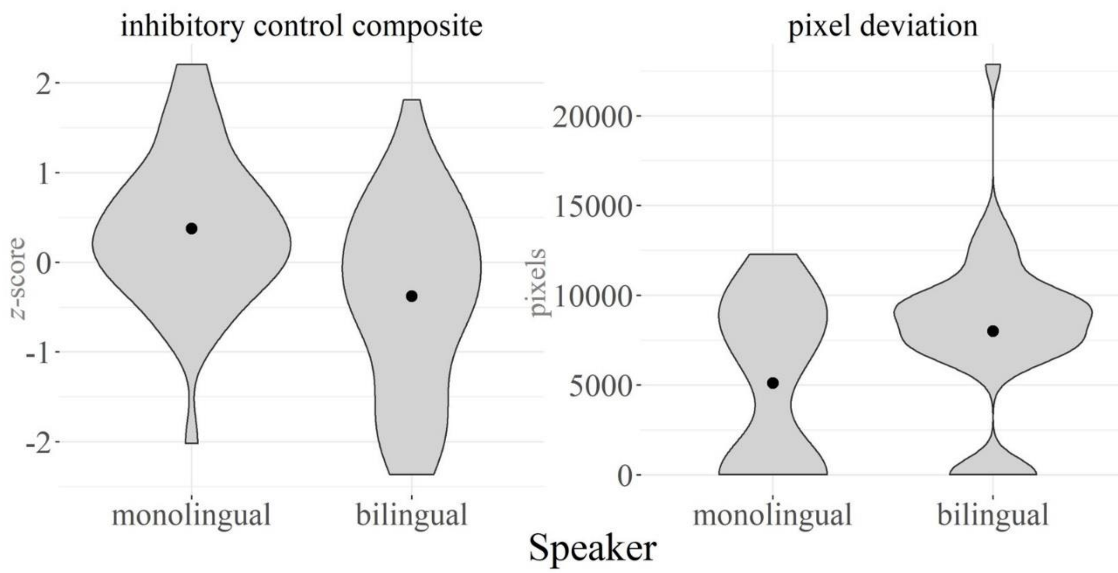

Figure 1: Distribution and mean (indicated by the black dot) of the inhibitory control composite score (left) and pixel deviation (right) per speaker group (*x*-axis). Higher inhibitory composite signifies better inhibitory control, and smaller pixel deviation signifies better fine-motor skills.

In order to better understand the finding that monolinguals demonstrated better domain-general inhibitory control and fine-motor abilities than bilinguals, exploratory analyses were conducted. One possibility was that effect is language driven, that is, two of the three inhibitory control tasks involved the L2 for the bilingual participants (i.e., Stroop and Antisaccade), which may have led to a disadvantage for that group. However, within bilinguals the only significant correlation between LexTALE and individual difference measures was with the reversed Stroop score ($r = -.49$, $p < 0.001$), suggesting that those with higher English proficiency exhibited poorer domain-general inhibitory control in the Stroop task. Further *t*-tests using individual task scores revealed that monolinguals still outperformed bilinguals in the Antisaccade task ($t = 4.28$, $p < 0.001$) and the Stop Signal task ($t = 2.03$, $p = 0.045$), and the two groups were not significantly different from one another in the Stroop task ($t = 0.58$, $p = 0.566$). Therefore, poorer inhibitory control composite scores do not appear to be due to any element of L2 use during the inhibitory control tasks.

In addition to pixel deviation, monolinguals also displayed better computer mouse fine-motor skills in terms of time on target ($t = 3.10$, $p = 0.003$). It is possible that bilinguals had overall less experience or familiarity using computers across their lifetime. This situation would result in overall slower reaction times during the inhibitory control tasks as well as the described results of the pursuit rotor task. Indeed, bilinguals were slower than monolinguals during the no-signal trials of the Stop Signal task ($t = -10.99$, $p < 0.001$) and in the neutral block of the Stroop task ($t = -13.48$, $p < 0.001$). Moreover, recall that bilinguals had spent approximately an average of four years in the United States. Although some monolinguals reported learning a second language, none reported any time in another country with the express purpose of language learning. Monolinguals appear to be more homogenous in their exposure to one culture, primarily that of the United States. Bilinguals were more heterogenous in this aspect, meaning that their exposure to computer technology may be more varied than that of monolinguals. This is one possible explanation, given that bilingual participants were on average older and had more years of education than the monolingual participants.

Quite unexpectedly, monolinguals consistently surpassed bilinguals in measures of domain-general inhibitory control and fine-motor skills. All three domain-general inhibitory control tasks were reliable and were combined in order to capture domain-general inhibitory control skill not specific to any one task, as has been recommended in the literature (Baum & Titone, 2014). The group difference in the inhibitory control composite score does not appear to be driven by proficiency or the fact that some tasks required L2 use. The pursuit rotor assessed fine-motor skills with a mouse, and in both measures of performance, monolinguals demonstrated better mouse control skills than bilinguals. This finding tempers the monolingual advantage in domain-general inhibitory control, as monolinguals were faster on all individual difference tasks. Therefore, one possibility is that the current group of monolinguals did possess greater domain-general inhibitory control than bilinguals, but not to the degree displayed, and another possibility is that both groups were similar in their domain-general inhibitory control abilities after computer performance skills were taken into account. In either scenario, the current study certainly did not reveal a cognitive advantage for late bilinguals. Next, we will examine the results of the visual-world task.

Visual world

In addition to the four visual-world outcomes submitted to multilevel models, mouse cursor positions were extracted for the purposes of visualization and the proportional proximity analysis in this section. Because each trial differed in length, I examined *x*- and *y*-coordinates at time normalized timestamps instead of raw time stamps. Time normalization applied a linear interpolation to create 101 time bins for each trial, regardless of length, to allow comparisons to be made across trials of different lengths (Freeman & Ambady, 2010). For visualization purposes, trajectories were also remapped so that the same side contained the target (i.e., left) and the competitor (i.e., right), space normalized (i.e., all trials began and end at the same points), and plotted with the mousetrap package (Kieslich, Wulff, Henninger, & Haslbeck, 2017).

Mouse-tracking studies do not typically trim data beyond incorrect responses, therefore neither did the current study (e.g., Spivey, Grosjean, & Knoblich, 2005). One reason may be that there are several outcomes, so either each is individually trimmed by, for instance, three standard deviations per person, or one outcome is chosen as the outcome by which to trim. Visual inspection of the individual trajectories for outliers—figures herein represent aggregated trajectories—lead to identification of 11 trials with raw AUC values greater than five. Most raw AUC values range between negative two and two, therefore I eliminated these 11 trials as outliers. The trials did not originate from a single participant or condition. Other exclusions included filler trials and incorrect experimental trials (2.14% or 59 trials).

Some previous mouse-tracking literature used *z*-score transformations of AUC and MD per participant across all trials (e.g., Farmer, Cargill, Hindy, Dale, & Spivey, 2007). As the raw AUC and MD values were not in meaningful units (e.g., pixels or a length unit, such as centimeters), I also used the *z*-score. Therefore, changes in these two outcomes in multilevel models can be interpreted in terms of standard deviations (e.g., bilinguals demonstrated MDs that were one standard deviation greater than monolinguals, meaning their mouse trajectories veered more towards the competitors).

Linear multilevel modeling was employed for logged RT, AUZ *z*-score, and MD *z*-score outcomes. Flips from the positive side of the *x*-axis to the negative represent counts, therefore logistic models were employed with a Poisson distribution specified. Raw RTs were not normally distributed and applying a log transformation increased the Shapiro-Wilk test statistic *W* from 0.92 to 0.98. Although significance did not change, I used the more normally distributed logged RT in the multilevel models because logged RTs better reflect psycholinguistic aspects of input processing. AUC and MD *z*-scores were not normally distributed and applying a log transformation did not increase normality.

Reliability was calculated per transformed outcome for each of the four lists. That is, each list was treated as a different form of the task since a given item changed conditions across lists, which would affect the outcomes (e.g., I predicted that RTs would slow down in the related condition). As seen in Table 6, reliability was relatively high, with the lowest being 0.68 for AUC in list D.

Table 6: Odd-even reliability of visual-world outcomes by list

	Outcome			
List	RT	AUC	MD	Flips
A	0.98	0.89	0.87	0.86
B	0.96	0.94	0.91	0.93
C	0.94	0.90	0.92	0.88
D	0.95	0.68	0.81	0.88

With this information, analyses were then conducted to inform the first research question: Does phonological overlap induce lexical competition?

RQ1

Descriptive statistics, such as means, standard deviations, and ranges, of each visual-world outcome, untransformed, are displayed in Table 7. Transformed outcomes were submitted to multilevel models with condition as the only fixed effect, with random intercepts for participants and items as well as by-participant random slopes for condition. Figure 2 displays the average mouse trajectories from the *Start* button to the target picture in each condition. Mouse trajectories are useful indicators of the AUC and MD outcomes.

Table 7: Visual-world untransformed outcome descriptive statistics per condition

	Unrelated		Related	
Outcome	Mean (SD)	Range	Mean (SD)	Range
RT	1173.28 (213.17)	[612 , 2050]	1199.81 (221.7)	[571 , 2164]
AUC	0.27 (0.74)	[-2.42 , 4.62]	0.48 (0.97)	[-0.97 , 4.89]
MD	0.15 (0.40)	[-0.92 , 1.65]	0.24 (0.47)	[-1.18 , 1.55]
Flips	6.98 (3.44)	[1 , 22]	6.88 (3.27)	[1 , 22]

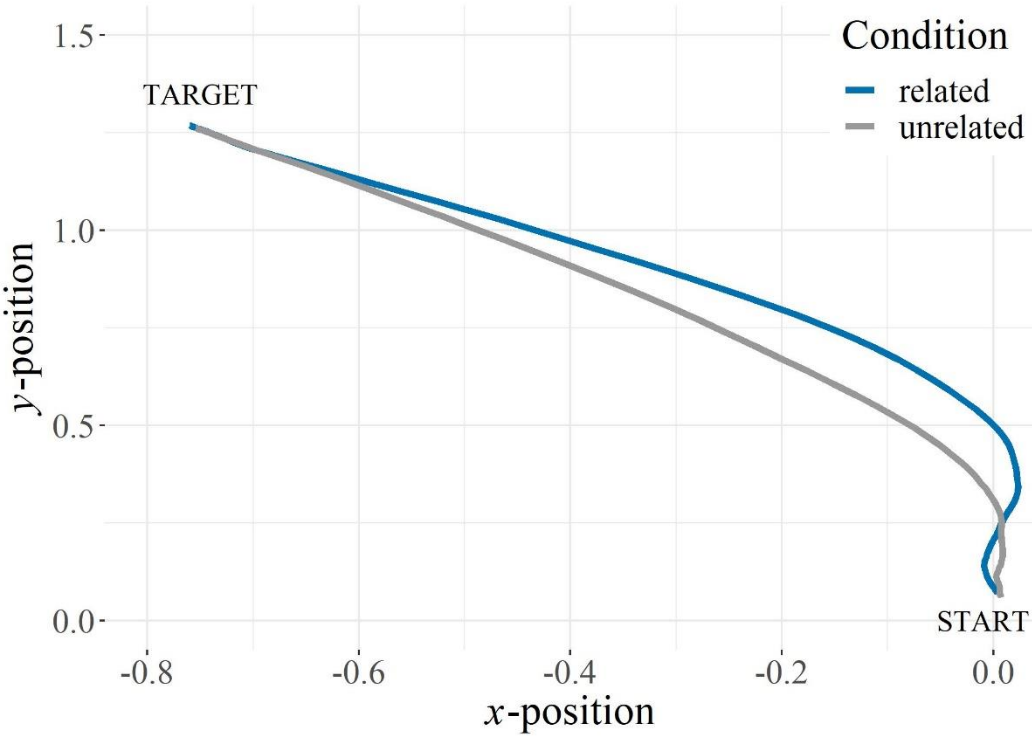

Figure 2: Average mouse trajectories per condition

For three out of four outcomes, the related trials showed evidence of lexical competition more than the unrelated trials. Table 8 displays the main effect of condition, the only fixed effect, for each outcome in the contrast coded models (for full model output, see Appendix F). Note that the test statistic for flips was Chi-square instead of *F* due to the logistic rather than linear multilevel model. The related trials displayed longer RTs and larger AUC and MD (RT: $z = -3.80$, $p < 0.001$; AUC: $z = -6.01$, $p < 0.001$; MD: $z = -5.26$, $p < 0.001$). In estimated model means, participants took 31.00 ms[1] longer to click on related trial targets, and their mouse trajectories veered more towards the competitor in related trials. The AUC *z*-score estimated model means signified that AUC

[1] For ease of interpretation, the logged RT estimated marginal means were transformed back to raw RTs by computing the exponential function for any significant category group differences. For continuous variables, the exponential function was imposed on the addition of the intercept and model estimate *b* to determine the slope.

were almost a quarter of a standard deviation larger in related than in unrelated trials (0.242), and a similar magnitude of difference was observed in the MD *z*-score outcome (0.221). The number of flips over the *x*-axis did not significantly differ between related and unrelated trials.

Table 8: Visual-world results per outcome of models with all speakers and trials

Outcome	*df*	*F*
RT	1, 81.39	13.25***
AUC	1, 31.80	35.70 ***
MD	1, 86.33	27.22 ***
Flips	1	0.77^

Notes. ****p* < 0.01, ***p* <.01, **p* <.05, +*p* <.1. ^ denotes a Chi-square statistic instead of *F*

Following this knowledge, all four outcomes were analyzed with regard to the third research question: Does lexical competition vary due to frequency, inhibitory control ability, speaker type, or their interactions?

RQ3

The visual-world variables relevant for the research question three investigations were condition (unrelated, related), speaker (monolingual, bilingual), target-competitor frequency difference (continuous), IC composite (continuous), and LexTALE score (continuous). The frequency difference variable was derived by subtracting the log frequency of the competitor from that of the target. It captured two pieces of information: 1) the polarity indicated if the competitor was higher or lower in frequency than the target (i.e., negative denotes a competitor that was higher frequency than the target), and 2) the value indicated the absolute difference in frequency between the target and competitor. Therefore, the model estimate for the fixed effect of frequency difference represents not only if the target frequency relative to the competitor frequency affected the outcomes (positive or negative estimate), but also to what degree the absolute log frequency difference affected the outcomes. LexTALE was only included in models with bilingual participants.

Descriptive statistics for each untransformed outcome are listed in Table 9. It was predetermined to explore the effects of speaker, frequency difference, IC composite, and LexTALE in related trials only. Frequency was designed to only influence related trials and the other variables were predicted to play a role in lexical competition, which was stronger in the related trials. Therefore Table 9 reflects outcomes in only the related trials.

Table 9: Visual-world descriptive statistics for related trials

Outcome	Monolingual mean (SD)	Bilingual mean (SD)	Frequency difference correlation	IC composite correlation	LexTALE correlation^
RT	1199.81 (221.7)	1367.71 (341.35)	-0.09**	-0.08**	-0.06
AUC	0.48 (0.97)	0.7 0 (1.15)	-0.03	-0.04*	-0.03
MD	0.24 (0.47)	0.36 (0.53)	-0.03	-0.04+	-0.03
Flips	6.88 (3.27)	6.95 (3.35)	<0.01	<0.01	<0.01

Notes. ****p* < 0.01, ***p* <.01, **p* <.05, +*p* <.1. ^LexTALE score correlations applied to bilingual participants only.

Each outcome was submitted to a multilevel model with fixed effects of speaker, frequency difference, IC composite, and their interactions with all speakers for related trials (i.e., LexTALE was not included at this stage). The results of each fixed effect parameter per outcome are listed in Table 10.

Table 10: Visual-world results per outcome of models with all speakers and related trials

	RT		AUC		MD		Flips	
Effect	*df*	*F*	*df*	*F*	*df*	*F*	*df*	*Chi-square*
Speaker	1 , 77.89	16.03***	1 , 37.89	0.87	1 , 41.79	0.15	1	0.02
IC composite	1 , 79.57	<0.01	1 , 411.17	0.19	1 , 326.7	<0.01	1	0.07
Frequency difference	1 , 1141.99	17.69***	1 , 1254.73	1.07	1 , 1268.48	1.09	1	1.53
Speaker x IC composite	1 , 81.84	0.12	1 , 314.34	0.23	1 , 192.39	0.12	1	0.77
Speaker x Frequency difference	1 , 1145.53	1.11	1 , 1256.06	0.20	1 , 1264.59	0.05	1	0.93
IC composite x Frequency difference	1 , 1143.76	2.05	1 , 1260.66	<0.01	1 , 1264.88	0.22	1	1.47
Speaker x IC composite x Frequency difference	1 , 1091.58	0.87	1 , 1258.41	0.48	1 , 1271.65	0.67	1	1.37

Notes. ***p < 0.01, **p <.01, *p <.05, +p <.1.

The only outcome with any significant parameters was the log transform of RTs. The contrast coded model indicated a main effect of speaker, and the estimated marginal means specify that bilinguals were 148.97 ms slower than monolinguals (z = -4.01, p = 0.001) for the average IC composite and frequency difference. Across speaker type, a larger and more positive difference in frequency between target and competitor led to faster RTs (b = -0.003, SE < 0.007, df < 0.01, t = -4.21, p < 0.001). When the competitor was log frequency 1.00 greater than the target (i.e., approximately 0.18 instances per million in the SUBTLEX-US corpus), this led to a 40.69 ms decrease across speakers and for the average IC composite score. Both results were predicted, but so was an interaction whereby bilinguals were more affected by the frequency difference, which did not occur.

Additionally, the IC composite score did not affect any outcome as a main effect or in any interactions.

To give a sense of the impact of relative competitor frequency on mouse trajectories, Figure 3 and Figure 4 display the average mouse trajectories for each speaker group. Since target-competitor frequency difference was continuous, the figures contain one trajectory per quartile, with blue indicating trials where the competitor was higher frequency than the target and green indicating trials where the competitor was lower frequency than the target.

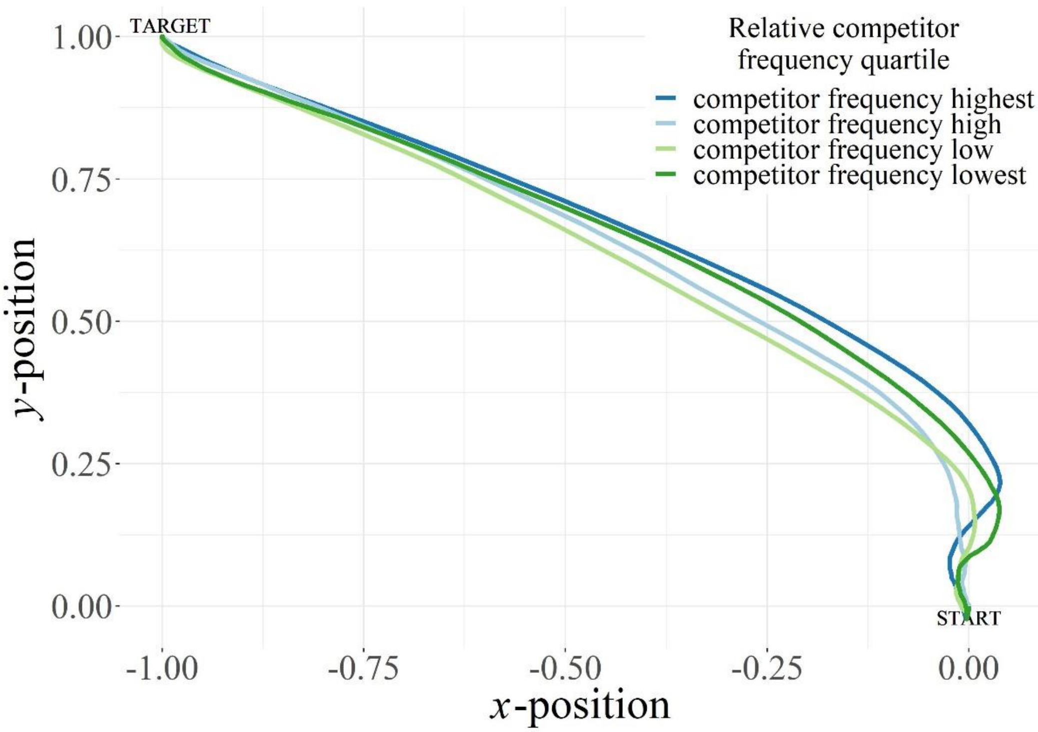

Figure 3: Monolingual average mouse trajectories by target-competitor frequency difference in related trials.

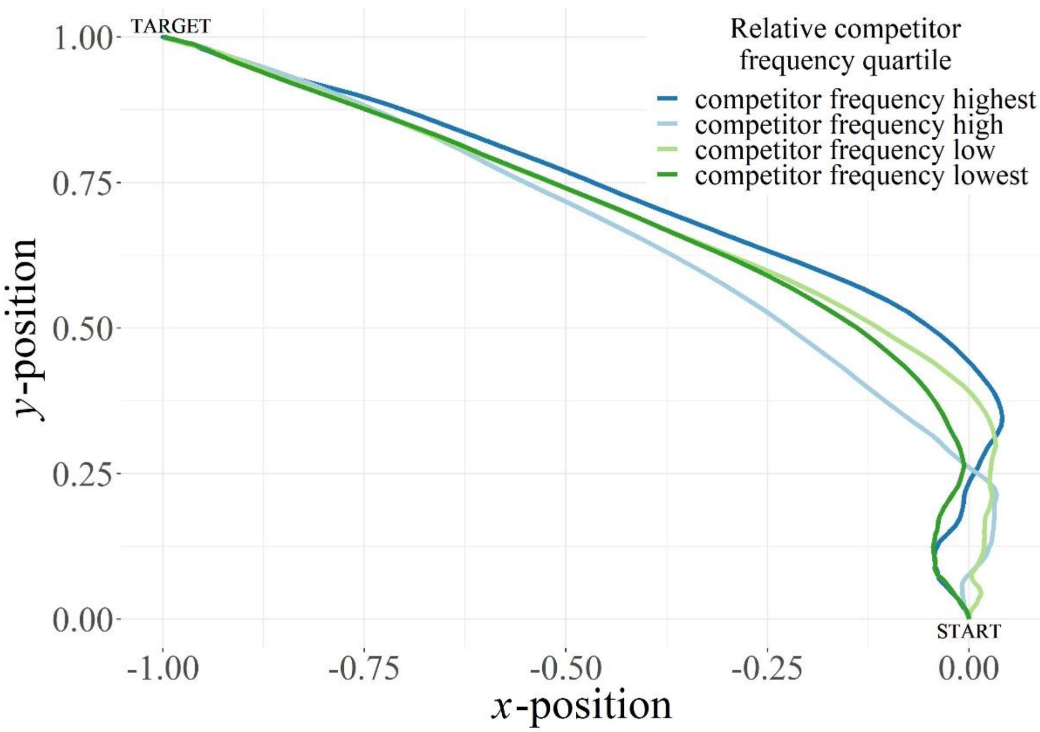

Figure 4: Bilingual average mouse trajectories by target-competitor frequency difference in related trials.

Although not a planned analysis, each inhibitory control task (i.e., Antisaccade, Stroop, and Stop Signal) was substituted for the IC composite in the previously described model to explore the possibility that the individual constructs underling the individual inhibitory control tasks could play a role. There remained no significant effects, except for a marginal interaction of the Stop Signal reaction time (SSRT) and speaker on AUC ($F(1, 1233.19) = 3.13$, $p = 0.080$). Better inhibition in the SSRT score signaled less distraction for bilinguals, but better SSRT scores signaled more distraction for monolinguals ($z = 1.77$, $p = 0.077$).

The effect of proficiency was explored by running very similar analyses but splitting by speaker group. Separate models for bilinguals and monolinguals enabled entering LexTALE into the bilingual models since no issues of collinearity would arise.

Models with only monolinguals and related trials included fixed effects of frequency difference, IC composite, and their interactions. Similar to the models with all speakers (Table 10), only log RT demonstrated any significant effects (see Table 11). The model estimated that monolinguals were 28.23 ms slower in trials with a competitor higher by 1.00 log frequency than the target for the monolingual average IC composite score (*b* = 0.024, *SE* = 0.010, *df* = 515.20, *t* = -2.48, *p* = 0.013).

Table 11: Visual-world results per outcome of models with only monolinguals and related trials

	RT		AUC		MD		Flips	
Effect	*df*	*F*	*df*	*F*	*df*	*F*	*df*	*Chi-square*
IC composite	1, 40.45	0.06	1, 391.75	<0.001	1, 213.34	0.04	1	0.83
Frequency difference	1, 515.17	6.16*	1, 667.38	0.16	1, 667.23	0.33	1	0.63
IC composite x Frequency difference	1, 518	0.28	1, 675.52	0.25	1, 675.81	0.65	1	0.12

Notes. ****p* < 0.01, ***p* <.01, **p* <.05, +*p* <.1.

Bilingual-only models included fixed effects of LexTALE, frequency difference, IC composite, and their interactions (see Table 12 for significance of each parameter). RT was consistent in demonstrating frequency difference effects — bilinguals were also slower when competitors were higher in frequency than targets at a rate of 49.12 ms per 1.00 log frequency difference for the average IC composite and LexTALE scores (*b* = -0.038, *SE* = 0.020, *df* = 257.30, *t* = -1.91, *p* = 0.058). For both speaker groups, the relative frequency of competitor and target impacted RTs to the target. Another way to view this result is that higher-frequency targets were processed faster in both L1 and L2 because (a) higher-frequency words were accessed faster, and also (b) lower-frequency competitors were not strong competitors. Thus, RT was primarily indicative of the speed of lexical access and (less directly) of lexical competition. The other mouse-tracking measures more directly assessed the impact of a lexical competitor.

In bilinguals, these remaining mouse-tracking outcomes signaled an effect of proficiency and frequency. Both AUC and MD indicated that higher proficiency led to lower AUC and MD, meaning mouse trajectories deviated less towards competitors in related trials for more proficient individuals (AUC: $b = -0.015$, $SE = 0.006$, $df = 191.50$, $t = -2.33$, $p = 0.021$; MD: $b = -0.012$, $SE = 0.006$, $df = 43.12$, $t = -1.88$, $p = 0.067$). For the average IC composite and LexTALE scores, AUC increased, signaling greater competitor distraction by 0.015 of the AUC standard deviation for every 1% decrease in LexTALE score. The main effect of LexTALE was marginal for MD, as was the main effect of frequency difference in AUC, which showed that competitors higher in frequency than targets had more mouse trajectory deviations toward the competitors ($b = -0.213$, $SE = 0.111$, $df = 430.40$, $t = -1.91$, $p = 0.056$). For every 1.00 log frequency decrease between target and competitor, AUC increased by 0.213 raw AUC standard deviations. AUC displayed effects of frequency in bilinguals that were not observed in monolinguals, albeit marginally, suggesting that higher-frequency competitors did receive more activation and were therefore stronger competitors for this speaker group. Proficiency had a stronger relationship with AUC and MD in bilinguals, suggesting that those with higher L2 proficiency were less affected by phonologically related competitors.

Table 12: Visual-world results per outcome of models with only bilinguals and related trials

	RT		AUC		MD		Flips	
Effect	*df*	*F*	*df*	*F*	*df*	*F*	*df*	*Chi-square*
IC composite	1, 37.98	0.07	1, 434.03	0.28	1, 53	0.02	1	0.92
Frequency difference	1, 257.35	3.64+	1, 430.35	3.67+	1, 53.57	2.61	1	0.56
LexTALE	1, 37.26	0.45	1, 191.51	5.42*	1, 43.12	3.55+	1	0.18
IC composite x Frequency difference	1, 375.18	0.65	1, 478.99	1.72	1, 51.01	1.4	1	0.1
IC composite x LexTALE	1, 36.25	0.36	1, 32.59	0.95	1, 29.86	0.8	1	1.03
Frequency difference x LexTALE	1, 413.87	<0.001	1, 487.14	1.99	1, 67.14	1.63	1	0.12
IC composite x Frequency difference x LexTALE	1, 498.42	0.52	1, 549.51	1.9	1, 50.46	2.25	1	3.00+

Notes. ****p* < 0.01, ***p* <.01, **p* <.05, +*p* <.1.

Flips, which had previously not shown any significant effects, had a marginal interaction of the IC composite, frequency, and LexTALE ($b = 0.006$, $SE = 0.003$, $z = 1.79$, $p = 0.074$). Figure 5 depicts this interaction by simplifying the continuous frequency difference into quartiles, as in previous figures, and displaying trend lines for only higher or lower IC composite scores. In trials with higher-frequency competitors (i.e., the stronger competitors), neither LexTALE nor IC were strongly influencing the odds of flipping towards the competitor, suggesting that neither factor aided in managing lexical competition when the competitor was strongest. As competitor frequency decreases relative to target frequency (i.e., competitor strength weakens), these variables did influence the odds of flipping across the *x*-axis, such that for participants with poorer IC ability, those who also had better proficiency were flipping less.

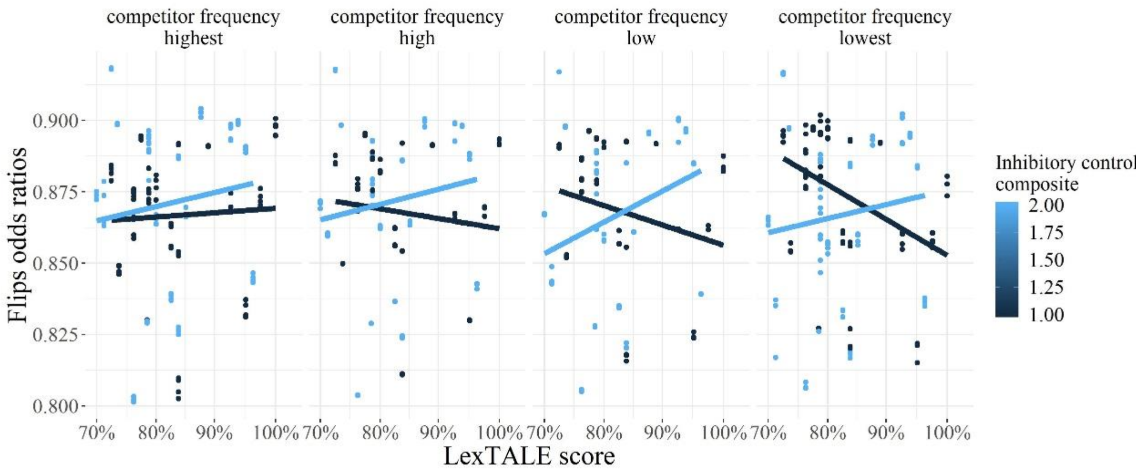

Figure 5: Flips odds ratios depicting the significant interaction of LexTALE, target frequency, and inhibitory control composite in the bilingual group, related trials only model

Exploratory analyses

The following section presents two analyses that were not originally planned (i.e., exploratory), but given that the results did not strongly reflect the predictions, these analyses could potentially add insight and are present in past studies examining similar topics. The first directly correlated lexical inhibition and domain-general inhibition for each visual-world outcome. The second investigated the proximity of the mouse cursor to the competitor or target in different conditions and speaker groups over the time course of a trial.

<u>Lexical inhibition scores</u>

Previous studies assessing the relationship between lexical and domain-general inhibition calculated correlations between lexical inhibition scores with domain-general inhibition scores (e.g., Blumenfeld & Marian, 2007). Since the planned analyses did not yield any effects of the IC composite, I took a similar approach to determine if a simple relationship between the two variables existed in the current study. Lexical inhibition scores per participant were calculated for each visual-world outcome by subtracting the average outcome on related trials from the average outcome on unrelated trials, first across frequency difference and then for each categorical frequency difference (i.e., higher-frequency competitor or lower-frequency competitor). Positive lexical inhibition scores indicated that targets on related trials were inhibited (i.e., slower or more deviations towards target), and negative lexical inhibition scores indicate that targets on related trials received facilitation from the presence of a phonologically related competitor. The lexical inhibition scores used log RT and AUC and MD *z*-scores.

Table 13 details the correlations among lexical and domain-general inhibition scores across target frequencies with *p*-value Holm corrections for multiple comparisons. The IC composite score did not have a strong nor significant correlation with any of the visual world outcome lexical inhibition scores. The positive values for the lexical inhibition averages corroborate the findings of the multilevel models — related trials induced greater lexical competition than unrelated trials.

Table 13: Correlations, means, and standard deviations of lexical and domain-general inhibition scores across target frequency

Measure	Mean (SD)	1	2	3	4
1. IC composite	0.01 (1.01)				
2. RT	0.02 (0.06)	-0.08			
3. AUC	0.24 (0.38)	0.03	0.37**		
4. MD	0.22 (0.39)	0.04	0.36**	0.95**	
5. Flips	0.04 (1.12)	-0.16	0.27*	0.13	0.15

Notes. ****p* < 0.01, ***p* <.01, **p* <.05, +*p* <.1.

Target frequency was included in the experimental design with the prediction that competitors higher in frequency than the target would be harder to inhibit due to its greater lexical activation levels. When lexical inhibition scores were split by competitor frequency (higher or lower categorically, see Table 14), there remained no significant correlations between lexical inhibition and the IC composite. The same correlational analyses were split by speaker type, but still no significant correlations emerged among lexical and domain-general inhibition scores.

Table 14: Correlations, means, and standard deviations of lexical and domain-general inhibition scores among items with higher- or lower-frequency competitors

Measure	Mean (SD)	1	2	3	4	5	6	7	8
1. IC composite	0.01 (1.01)								
2. RT LF	0.03 (0.09)	-0.18							
3. RT HF	0.02 (0.08)	0.07	0.14						
4. AUC LF	0.26 (0.49)	0.10	0.21+	0.19+					
5. AUC HF	0.22 (0.49)	-0.04	0.20+	0.24*	0.18				
6. MD LF	0.24 (0.51)	0.14	0.18	0.16	0.95**	0.15			
7. MD HF	0.20 (0.53)	-0.07	0.23*	0.21+	0.15	0.92**	0.10		
8. Flips LF	-0.01 (1.45)	-0.20+	0.23*	-0.10	0.08	0.04	0.08	0.07	
9. Flips HF	0.11 (1.51)	-0.04	0.14	0.32**	0.05	0.13	-0.03	0.21+	0.15

Note: *** denotes $p < 0.01$, ** denotes $p <.01$, * denotes $p <.05$, + denotes $p <.1$. HF refers to items containing higher-frequency competitors, and LF refers to items containing lower-frequency competitors.

Proportional proximity

Although mouse tracking and eye tracking are different methodologies, both allow for investigation into the word recognition process throughout a trial, which is an advantage over traditional button-press RTs. The four outcomes previously analyzed summarize the competitor influence over an entire trial, and proportional proximity should provide converging information and new information about when during the trial competitor influence was the strongest. Because each trial in the visual-world task differed in length, time-normalized timestamps were used instead of raw-time timestamps. It was possible to use both time-normalized and raw-time coordinates for such an analysis, although previous work on the time course using raw time data limited the investigation to the first second of a trial to avoid any confound of different trial lengths (e.g., Incera & McLennan, 2016). The advantage of time-normalized data was that the entire trial was examined, but the advantage of using raw time data was that researchers can know the millisecond when divergence occurred, which was not possible with time-normalized data. The MouseTracker analysis software created 101 timestamps for each trial, therefore each timestamp can be thought of as a percentage of the trial time that has elapsed. For instance, at the 75th normalized timestamp approximately 75% of the trial time has passed.

Eye-tracking typically presents fixations to a target or competitor across time. In mouse-tracking, a similar measure is the proportional proximity of a mouse cursor to either the target or competitor at each normalized timestamp. Proportional proximity of a mouse trajectory coordinate (i.e., *x*- and *y*-position) to either the target or competitor first required calculation of the Euclidean distance. This metric combined velocity and spatial attraction from the *x*- and *y*-coordinates at each normalized timestamp.

Euclidean distance was calculated using the following formula (Hehman, Stolier, & Freeman, 2015):

$$\text{distance}((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2}$$

The *x*- and *y*-values represented the *x*- and *y*-coordinates at each time-normalized timestamp, and (a, b) represented the coordinates of the center of the target or competitor. Euclidean distance was calculated for each timestamp twice – once for distance to the target and once for distance to the competitor.

Previous mouse-tracking literature transformed Euclidean distance into proportional proximity using the following formula (Spivey et al., 2005):

$$1 - \text{distance} / \max(\text{distance})$$

Therefore I additionally calculated proximity to the target and competitor. A larger target proximity value, for instance, indicated that the mouse cursor was closer to the target at that timestamp for that participant and trial.

Proportional proximities were averaged across participants to produce the visualizations in this section and for inferential statistics. Following conventions, *t*-tests were performed at each timestamp, with simulations and collected data indicating that eight consecutive significant differences in coordinates, Euclidean distance, or proportional proximity represent a reliable divergence in trajectories (Dale, Kehoe, & Spivey, 2007). Preceding studies have reported significance with an alpha level of 0.05

(e.g., Spivey et al., 2005), thus *t*-tests results reported uncorrected *p*-value, but *p*-values with a Holm correction for multiple comparisons were included as well.

The average trial length was 1263.81 ms from the time the participant clicked in the *Start* button to the time they clicked on the target, which included a 500 ms mouse cursor freeze after clicking the *Start* button. After this freeze participants heard the target word. Thus the 50th timestamp on average occurred 631.90 ms after the *Start* button was clicked and the 75th timestamp at 947.86 ms.

The first research question asked about the difference between mouse trajectories in the unrelated and related conditions in the visual world task. That is, did mouse trajectories in each condition differ in their proximity to the target or competitor. Figure 6 displays the average proportional proximity across speakers to the target and competitor for each condition. Results of *t*-tests at each timestamp comparing average related to unrelated proportional proximity can be viewed in Appendix G.

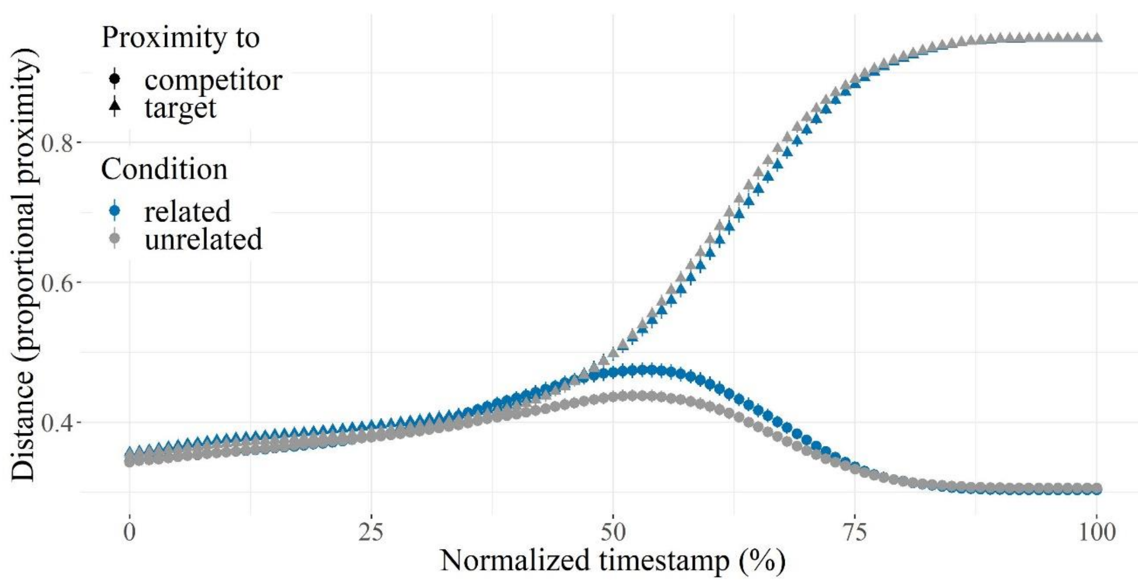

Figure 6: Proportional proximity to the target (triangles) or competitor (circles) in related (blue) or unrelated (grey) conditions averaged across speaker type. Bars indicate 95% confidence limits calculated via nonparametric bootstrap.

Although difficult to discern in Figure 6, *t*-tests indicate that participants' mouse cursor positions were closer to the target in the related condition than in the unrelated condition from the fourth to the 28th timestamp. In the later part of trials (i.e., 57th to 75th timestamp), mouse cursor positions in the unrelated condition were significantly closer to the target than those in the related condition. Therefore, when the target word was more likely to be heard (i.e., the later part of the trial), participants' mouse trajectories were closer to the target in trials with an unrelated competitor. Proportional proximity was closer (i.e., larger) to the competitor in the 35th to 72nd timestamp in related conditions compared to unrelated, suggesting that in the middle portion of related trials, participants' mouse trajectories were closer to the competitor than in unrelated trials.

The third research question investigated the difference in competitor frequency for the different speaker groups (i.e., monolingual and bilinguals). To examine this question using proportional proximity, I compared higher- and lower-frequency competitors in each speaker group in related trials. Although relative target-competitor frequency was continuous in the multilevel models, *t*-tests provided group comparisons, therefore the previously continuous variable was categorized into competitors of higher frequency (i.e., predicted to be stronger competitors) or lower frequency (i.e., predicted to be weaker competitors) than the target.

Figure 7 presents the monolingual proportional proximity to target and competitor pictures differentiated by relative competitor frequency (i.e., higher or lower than the target). Visually, in the later part of the trial after the target was heard, the lower-frequency competitors did not appear to be exerting any more influence on mouse trajectories than the higher-frequency competitors, either in proximity to the target or competitor. The *t*-tests revealed no significant differences between higher- and lower-frequency competitors in proportional proximity to the target. Monolingual mouse cursors were significantly closer to the lower-frequency competitors in the 20th through 23rd timestamp, although this pattern does not meet the eight consecutive timestamp criterion established by Dale et al. (2007) as a reliable trajectory divergence. Relative competitor strength, therefore, did not greatly influence lexical competitor as assessed by mouse trajectories during word recognition for monolinguals

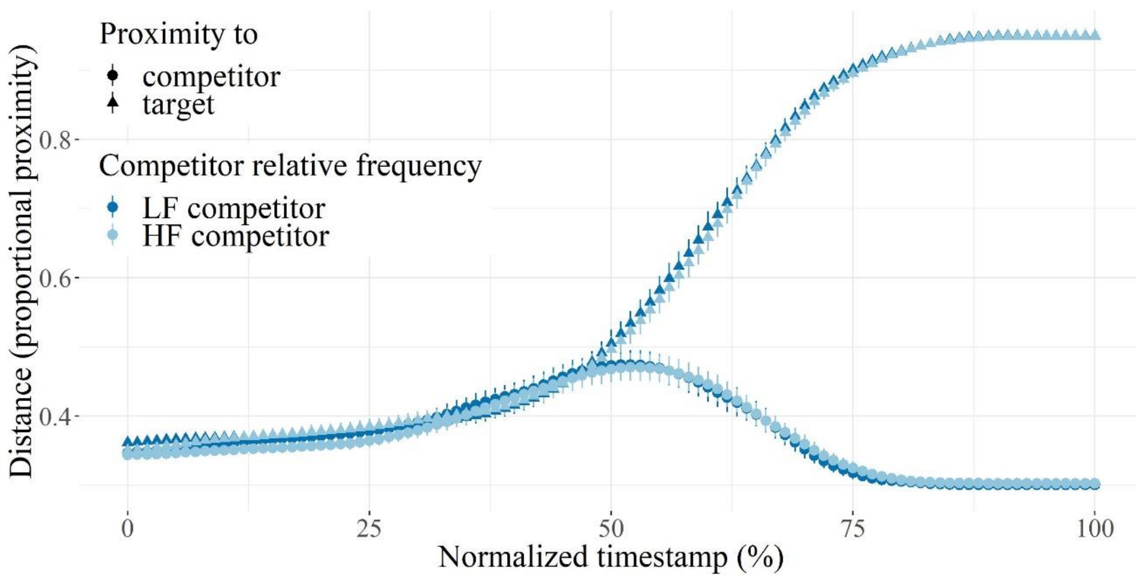

Figure 7: Monolingual proportional proximity to the target (triangles) or competitor (circles) in lower-frequency competitor (dark blue) or higher-frequency competitor (light blue) related trials averaged across participant. Bars indicate 95% confidence limits calculated via nonparametric bootstrap.

Figure 8 presents the same information as Figure 7 but for bilingual participants. Visually it appears that bilingual mouse cursers were closer to the target when the competitor was lower in frequency (i.e., a weaker competitor) from approximately the 50th to 75th timestamp. The *t*-tests demonstrated greater proportional proximity to the target for lower-frequency competitor trials for the 56th through 61st timestamp, again not meeting the Dale et al. (2007) criterion. Turning to proximity to the competitor, Figure 8 displays a situation whereby bilingual mouse cursors were on average closer to the competitor when the competitor was higher in frequency from approximately the 50th to 75th timestamp. However, lower- and higher-competitor proportional proximity to the competitor did not significantly differ at any timestamp. It is worth noting that there were wider confidence intervals on the bilingual averages per condition and timestamp, perhaps indicating a larger degree of variation. It is this variation that may have led to the non-significant differences in proportional proximity to the competitor.

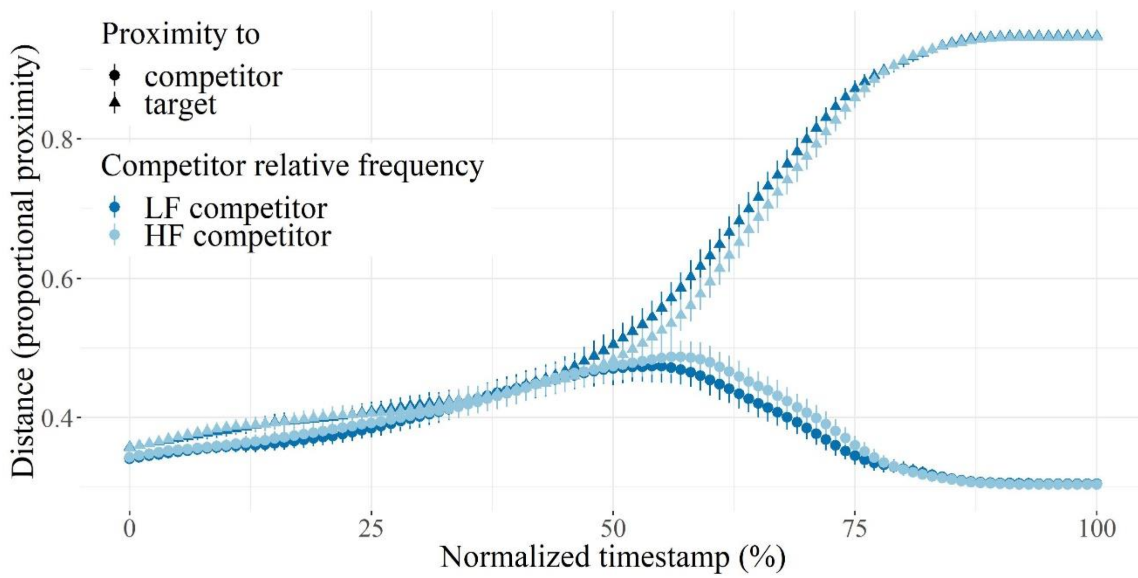

Figure 8: Bilingual proportional proximity to the target (triangles) or competitor (circles) in lower-frequency competitor (dark blue) or higher-frequency competitor (light blue) related trials averaged across participant. Bars indicate 95% confidence limits calculated via nonparametric bootstrap.

Overall in the visual world task, there was evidence of lexical competition in the finding that related trials were slower and had more mouse trajectory deviations towards competitors than unrelated trials. Models with both speakers with only related trials revealed a consistent effect of speaker and frequency on RTs: bilinguals were slower as were higher-frequency competitor trials. When speakers were divided into separate models in order to examine the effects of the bilingual-specific variable, proficiency, monolinguals demonstrated a similar frequency effect in RTs only. Bilinguals, on the other hand, displayed a frequency effect not only in RTs but a marginal one in AUC, as well as effects of proficiency. In the proportional proximity analysis, visually but not statistically it appeared that bilinguals again were more influenced by competitor frequency. Since these models were limited to related trials, one conclusion is that language proficiency and frequency influenced mouse trajectories in bilinguals more strongly than in monolinguals. Next, we will examine the same variables in the phonological priming task.

Phonological priming

In the phonological priming task RT data, 6.66% (221) of trials were excluded due to incorrect responses (e.g., indicating a real word after hearing a nonce word). Across all trials, filler and experimental, the lowest accuracy of any one participant was 76%, well above the 50% chance level. Neither the raw RTs nor a log transformation was normally distributed, but log transformation did improve the long, positive tail. RTs were overall very fast in this task (M = 362.79 ms, SD = 233.46 ms). Trimming by three standard deviations per person did not remove very short responses (e.g., 0.132 ms), which are usually desirable to remove because they were likely to be caused by error (i.e., too short to be informative for the task). Therefore, instead the data were trimmed by removing RTs below 50 ms and above 2,000 ms, which resulted in further exclusion of 3.15% (91 trials) of the priming data.

Reliability of RTs was calculated for each of the two lists. As in the visual-world task, each list was treated as a different form of the task since a given item changed conditions across lists. Odd-even reliability for list A was 0.88 and 0.77 for list B. After preprocessing, analyses were then conducted to inform the first research question: Does phonological overlap induce lexical competition?

RQ1

The average, untransformed RT in the related condition was 370.96 ms (*SD* = 205.06) and 367.05 ms (*SD* = 205.75) in the unrelated condition, and RTs were log transformed for all reported models. Condition did not significantly impact RTs in the priming task model with all speakers and condition as the only fixed effect (F (1, 81.05) = 0.14, p = 0.710). The average model estimated RT was 297.52 ms (*SE* = 18.51) for the unrelated condition and 300.02 ms (*SE* = 18.81) in the related condition, and they were not significantly different from one another (z = 0.38, p = 0.705). The related condition was predicted to demonstrate a decrease in RTs. While the average of both the raw and estimated model RTs per condition did not indicate a priming effect, there did appear to be participant and item variation in the degree of priming.

Following the procedure in Broersma (2012), Figure 9 displays the priming effect of each participant and item in a histogram form. Since relatedness of an item differed by list (e.g., item 21 was presented in an unrelated pair for participant 1, a bilingual, in list A and in a related pair for participant 2 in list B), the priming effect per participant and item was calculated by subtracting the RT on a related item from the speaker group average RT for the unrelated item (e.g., bilingual item 21 unrelated average RT – person 1, item

21 RT). Positive values along the *x*-axis in Figure 9 represent facilitation, and negative values represent inhibition. While the peak for both speaker groups is around 0 ms, indicating neither priming nor facilitation, there is considerable variability, especially in the bilinguals. The analyses for this task that apply to research question three attempt to explain this variation.

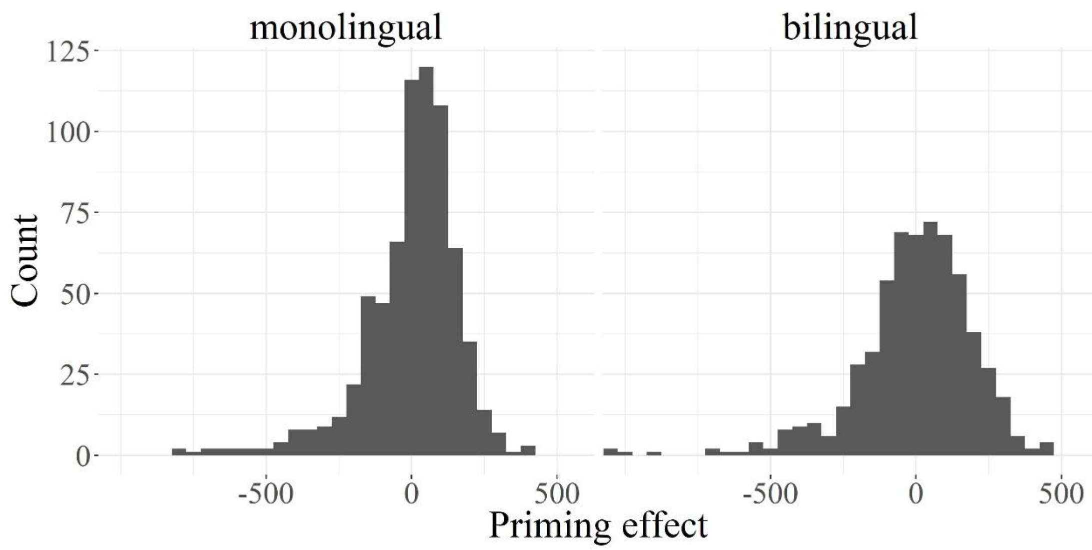

Figure 9: Histogram of priming effect per participant and item. Binwidth is 50 ms

RQ3

The priming experimental design variables related to the third research question were condition (unrelated, related), speaker (monolingual, bilingual), target frequency (continuous), IC composite (continuous), and LexTALE score (continuous). As proposed, some analyses contained only higher- or lower-frequency target trials, which was determined by a median of 3.16 log frequency in the experimental target words. As in the visual-world analyses, LexTALE was included with bilingual speakers only. The average RT for monolinguals was 332.17 ms (*SD* = 178.78, range = [50.08, 1552.49]) and 411.76 ms for bilinguals (*SD* = 224.93, range = [50.72, 1966.64]). RTs across all conditions and speakers correlated significantly with the IC composite score (r = -0.10, p < 0.01) and marginally with target frequency (r = -0.03, p = 0.08). Bilingual RTs across all conditions correlated significantly with LexTALE score (r = -0.12, p < 0.01).

It was predicted that bilinguals would show inhibition only in higher-frequency trials. If so, those lower-frequency trials would not be included in models with inhibitory control as there would be no reason to expect a connection between domain-general and lexical inhibitory control in trials with no lexical inhibition. A model with condition, speaker, and target frequency was conducted, and the results are presented in Table 15.

Table 15: Phonological priming results of a model with all speakers and all target frequencies

Effect	*df*	*F*
Condition	1, 117.43	0.16
Target frequency	1, 38.43	0.71
Speaker	1, 81.52	26.54***
Condition x Target frequency	1, 113.98	0.02
Condition x Speaker	1, 111.16	0.39
Target frequency x Speaker	1, 40.16	2.14
Condition x Target frequency x Speaker	1, 91.52	7.14**

Notes. ****p* < 0.01, ***p* <.01, **p* <.05, +*p* <.1.

A significant interaction of condition, speaker, and target frequency would signal that bilingual speakers treated lower-frequency target trials differently than monolingual speakers. This interaction was significant (b = 0.054, SE = 0.020, df = 91.52, t = 2.67, p = 0.009), in addition to a main effect of speaker. Bilingual estimated model means were 76.48 ms slower than monolingual (z = 5.127, p < 0.001) across condition and for the average (centered) target frequency. Figure 10 displays the interaction for the model predicted log RTs. Bilinguals demonstrated facilitation in the lower frequency range (i.e., related trials were faster than unrelated) and inhibition in the higher target frequency trials (i.e., related trials were slower than unrelated). Monolinguals exhibited the traditional inhibitory phonological priming effect in the lower frequency target trials but a flip to facilitation in the higher target frequency trials.

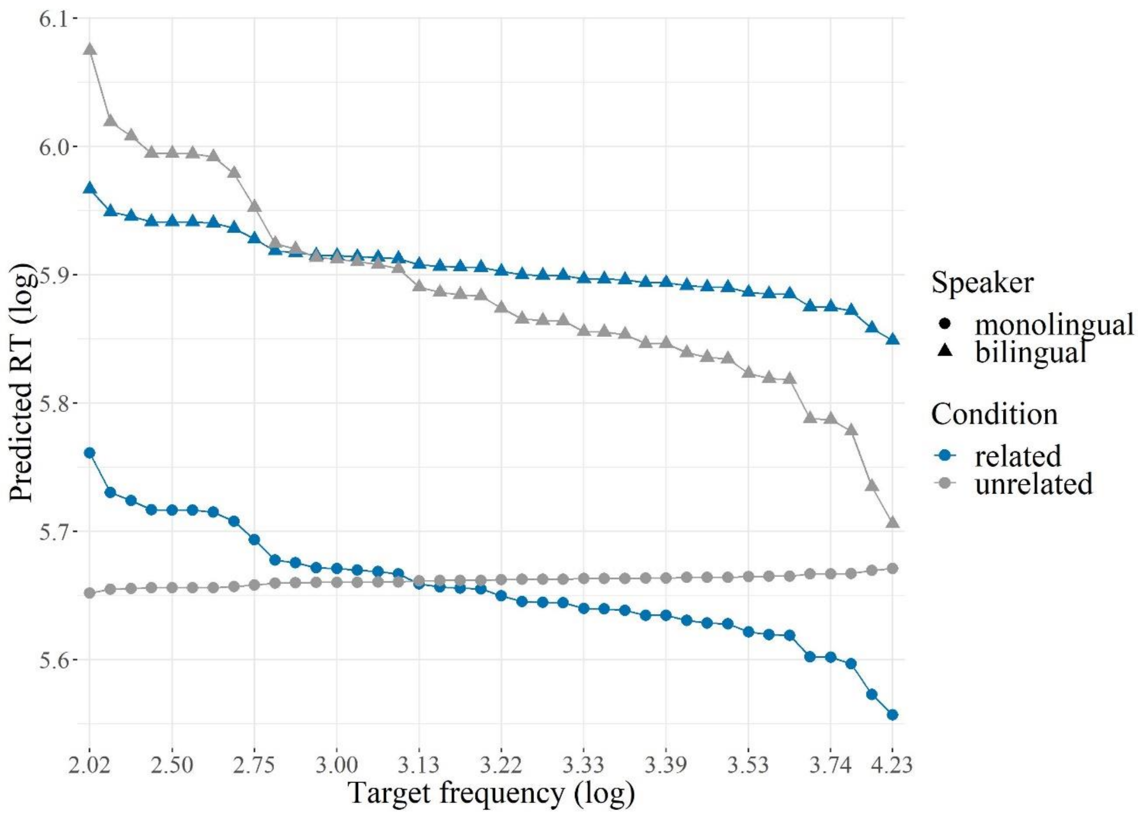

Figure 10: Phonological priming interaction of speaker, target frequency (log), and condition on model predicted log RTs

To give a sense of the magnitude of facilitation and inhibition, estimated model means were calculated for the most and least frequent target for each speaker group. In the lowest target frequency trial (i.e., 2.021 log frequency), bilinguals were 44.51 ms faster in the related trials than in unrelated trials, although not significantly (z = -1.45, p = 0.148), and monolinguals were 32.86 ms slower in the related trials (z = 1.62, p = 0.105). In the highest target frequency trial (i.e., 4.230 log frequency), bilinguals were 46.20 ms slower in the related trial (z = 2.04, p = 0.042), and monolinguals were 31.27 ms faster in the related trials (z = -1.72, p = 0.086), albeit marginally. Therefore, it appears this interaction was driven by the flip in the higher-frequency trials from inhibition in bilinguals to (marginal) facilitation in monolinguals. It is interesting to note that while not significant, facilitation in the lower-frequency trials was the numerical trend in bilinguals, as was predicted and seen in groups of English-Russian learners in Cook and Gor (2015).

Since it appears that bilinguals were exhibiting facilitation to some degree for lower-frequency trials, these items were removed for a model including the IC composite as well as speaker and condition. The logic was that if lower-frequency items did not produce lexical inhibition in bilinguals, then there would be no reason to recruit domain-general inhibitory control. Similar to most outcomes thus far, there was a main effect of speaker as well as an interaction between condition and speaker (see Table 16). Bilinguals' estimated model means were slower than monolinguals by 65.81 ms (z = 3.92, $p < 0.001$) across conditions and for the average IC composite score. The interaction estimated model means again indicated that bilinguals were significantly slower in the unrelated condition than in the related trial by 30.22 ms (z = 2.14, p = 0.032). That is, bilinguals displayed inhibition in these higher-frequency trials. Monolinguals were faster in the unrelated trials compared to related by 11.16 ms, although not significantly (z = -1.00, p = 0.317).

Table 16: Phonological priming results of a model with all speakers and higher-frequency trials

Effect	*df*	*F*
Condition	1, 560.76	0.67
IC composite	1, 53.25	2.36
Speaker	1, 65.05	15.42***
Condition x IC composite	1, 536.65	0.71
Condition x Speaker	1, 548.22	5.14*
IC composite x Speaker	1, 67.52	0.29
Condition x IC composite x Speaker	1, 531.4	1.96

Notes. ***p < 0.01, **p <.01, *p <.05, +p <.1.

The inhibitory control composite did not impact RTs in the phonological priming task, even in the trials predicted to induce the greatest amount of lexical inhibition. In an exploratory fashion, individual inhibitory control tasks (i.e., Antisaccade, Stroop, and Stop Signal) were substituted for the IC composite in the model described in Table 16. For all three models the main effects of speaker and interactions of condition and speaker remained.

Antisaccade, like the IC composite score, did not significantly impact RTs as a main effect ($F(1, 55.03) = 0.96$, $p = 0.330$) nor did Stroop scores ($F(1, 66.03) = 0.20$, $p = 0.660$). Stop Signal SSRT scores were significant as a main effect ($F(1, 75.43) = 5.86$, $p = 0.020$). Recall that this score was reversed, meaning that higher (i.e., better) SSRT scores led to faster RTs ($b = -0.001$, $SE = 0.001$, $df = 75.40$, $t = -2.42$, $p = 0.018$) at a rate of a 0.13 ms decrease in RT for every 1 ms increase in SSRT above the mean across speaker and condition.

Following the same procedure as the visual-world analyses, separate models were conducted for each speaker group in order to determine the effect of LexTALE on bilinguals alone. In this task, the decision to split also enabled the effect of inhibitory control to be examined in all frequency trials for monolinguals. In the monolingual only model, which included fixed effects of condition, frequency, the IC composite and their interactions, the IC composite did not influence RTs, and neither did condition nor target frequency (see Table 17). Accordingly, the speaker by condition and speaker by condition by target frequency interactions previously described can be interpreted to be driven by some priming effects among the bilinguals that the monolinguals did not exhibit in isolation.

Table 17: Phonological priming results of a model with monolinguals and all frequency trials

Effect	*df*	*F*
Condition	1, 49.22	0.24
Target frequency	1, 39.13	0.29
IC composite	1, 40.51	2.61
Condition x Target frequency	1, 49.92	1.97
Condition x IC composite	1, 47.26	1.63
Target frequency x IC composite	1, 163.19	0.76
Condition x Target frequency x IC composite	1, 43.74	0.17

Notes. ***p < 0.01, **p <.01, *p <.05, +p <.1.

Bilinguals alone were included in three multilevel models. The first contained both higher- and lower-frequency trials but not the IC composite due to the inclusion of the lower-frequency trials. The fixed effects were condition, LexTALE score, target frequency, and their interactions. As detailed in Table 18, there were significant main effects of LexTALE score and a marginal interaction of condition and target frequency.

Table 18: Phonological priming results of a model with bilinguals and all frequency trials

Effect	*df*	*F*
Condition	1, 65.52	<0.01
LexTALE	1, 37.47	5.00*
Target frequency	1, 38.34	0.92
Condition x LexTALE	1, 65.36	0.40
Condition x Target frequency	1, 649.45	3.61+
LexTALE x Target frequency	1, 78.79	0.55
Condition x LexTALE x Target frequency	1, 639.34	0.25

Notes. ***p < 0.01, **p <.01, *p <.05, +p <.1.

Bilingual participants with higher proficiency were faster (b = -0.009, SE = 0.004, df = 37.50, t = -2.24, p = 0.031) at a rate of a 3.11 ms decrease in RT for every 1% increase in LexTALE score, across conditions and for the average target frequency. Therefore, for a 10% increase in LexTALE score, bilingual participants were predicted to decrease RTs by 29.81 ms, which provides a slightly different interpretation due to the nature of the logarithmic scale. The interaction of condition and target frequency was marginal (*b* = -0.073, *SE* = 0.039, *df* = 649.40, *t* = 1.90, *p* = 0.058). Similar to the pattern in Figure 10, bilinguals exhibited facilitation in the lower frequency range and inhibition in the higher frequency range. The inclusion of LexTALE led to the knowledge that bilinguals RTs were impacted by proficiency, a proxy for entrenchment, but it did not interact with the other effects already described.

In the second bilingual-only model, the fixed effects were condition, LexTALE score, and the IC composite score. This model included only higher-frequency trials in order to determine the effect domain-general inhibitory control for the trials where lexical competition occurred (see Table 19). Bilinguals were marginally slower in the related condition compared to unrelated by 29.89 ms, according to the estimated model means (*z* = -2.35, *p* = 0.019) for the average LexTALE score. The marginal effect of LexTALE itself was observed as a decrease in RT by 2.94 ms for every increase in 1% LexTALE score across conditions (*b* = -0.009, *SE* = 0.005, *df* = 35.45, *t* = 1.90, *p* = 0.058). A 10% increase in LexTALE, then, led to RTs 28.21 ms faster. This model confirmed inhibition in bilinguals for the higher-frequency items and that increased proficiency led to marginally faster RTs.

Table 19: Phonological priming results of a model with bilinguals and higher-frequency trials

Effect	*df*	*F*
Condition	1, 276	3.56+
LexTALE	1, 35.45	3.79+
IC composite	1, 34.28	1.83
Condition x LexTALE	1, 264.33	0.14
Condition x IC composite	1, 250.75	0.59
LexTALE x IC composite	1, 34.99	0.34
Condition x LexTALE x IC composite	1, 254.68	0.5

Notes. ****p* < 0.01, ***p* <.01, **p* <.05, +*p* <.1.

The final bilingual-only model examined only lower-frequency trials and contained fixed effects of condition, LexTALE, and their interaction (see Table 20). Again, LexTALE was significant (b = -0.011, SE = 0.005, df = 37.50, t = -2.39, p = 0.022), with a 3.75 ms decrease in RT for every 1% increase in LexTALE score across conditions. A 10% LexTALE increase led to a 35.70 RT decrease. The following section describes the correlational analysis between lexical and domain-general inhibitory control in the priming task.

Table 20: Phonological priming results of a model with bilinguals and lower-frequency trials

Effect	df	F
Condition	1, 453.93	1.75
LexTALE	1, 37.47	5.69*
Condition x LexTALE	1, 453.1	1.08

Notes. ****p* < 0.01, ***p* <.01, **p* <.05, +*p* <.1.

Exploratory analyses

Mirroring the visual-world analyses, lexical inhibition (i.e., priming effect) scores per participant were calculated by subtracting the average RT on related trials from the average RT on unrelated trials per participant. Positive lexical inhibition scores indicate that targets on related trials were inhibited (i.e., slower RTs on related trials), and negative lexical inhibition scores indicate that targets on related trials received facilitation from the presence of a phonologically related prime (i.e., faster RTs on related trials). The correlation between phonological priming lexical and domain-general inhibition scores was not significant (r = 0.09, p = 0.380), nor were the same correlations when separated by speaker group (monolinguals: r = -0.17, p = 0.260; bilinguals: r = -0.05, p = 0.770).

Phonological priming lexical inhibition scores were correlated with the visual-world lexical inhibition scores to determine if there was a similar pattern of inhibition or facilitation within participants across tasks. The correlations were not significant nor strong across speaker groups (see the second column Table 21) nor when separated into monolingual and bilingual speakers (i.e., the two right columns).

Table 21: Correlations of phonological priming lexical inhibition scores with visual-world lexical inhibition scores

Visual-world lexical inhibition score	Correlation with priming lexical inhibition score All speakers	Correlation with priming lexical inhibition score Monolinguals	Correlation with priming lexical inhibition score Bilinguals
RT	0.04	0.11	-0.02
AUC	0.09	0.05	0.1
MD	0.03	-0.05	0.08
Flips	-0.19+	-0.19	-0.24

Notes. ****p* < 0.01, ***p* <.01, **p* <.05, +*p* <.1.

In conclusion, the predicted pattern for the priming task was that related trials would be slower than unrelated trials (i.e., inhibition) for most participants and for most items, except for lower-frequency items for bilinguals that were predicted to show facilitation. However, monolinguals demonstrated only a hint of the traditional priming effect in lower-frequency items. Although small, bilinguals displayed the predicted inhibitory effect in higher-frequency trials and a facilitatory effect in lower-frequency trials. In the model with only monolinguals, their RTs were not influenced by inhibitory control, condition, nor target frequency. In bilinguals, there was also no influence of the IC composite, but RTs were impacted by proficiency and target frequency. The participants with higher entrenchment were overall faster. Bilinguals demonstrated inhibition in higher-frequency trials, and some evidence of facilitation in lower-frequency trials.

Chapter 5: Discussion

The current study compared lexical competition management in two groups of speakers, late bilinguals and English-speaking monolinguals. The entrenchment account (Diependaele et al., 2013) suggests that bilinguals experience more lexical competition than monolinguals due to lexical representations that are less entrenched. Less entrenched lexical representations show weaker activation, especially in lower-frequency words, and lower phonolexical quality. As a result, bilinguals are overall slower at target word selection and more sensitive to lexical frequency than monolinguals. Weaker representations can even go so far as to lead to sublexical facilitation in the priming tasks. Given that the strength of representations increases with proficiency and is greater for high-frequency words, proficiency can serve as a proxy for the strength of an L2 representation, as does lexical frequency.

The bilingual advantage literature, on the other hand, predicts that bilinguals are more efficient due to experience with an increased number of lexical representations across two languages of a bilingual, and thus an increased number of competitors, during recognition (Bialystok, 1999). Domain-general inhibitory control was investigated as a compensatory mechanism during lexical competition in the current study, a link proposed by the bilingual advantage literature but consistent with both it and the entrenchment account. That is, those with better domain-general inhibitory control should better manage competition, whether that be bilinguals or monolinguals with larger vocabularies.

In the visual-world task, lexical competition was created by presenting a cohort competitor during target selection (e.g., *butter-bubble*). On the other hand, the priming task examined the degree of lexical inhibition of the target due to being one of many cohort competitors during prime recognition (e.g., *remain-remind*), assuming that slower target reaction time was a result of greater lexical inhibition. As such, lexical competition was overall more robustly observed in the visual-world task compared to the priming task. In line with the predictions of the entrenchment account, bilinguals were more sensitive to the relative frequency of target and competitor in the visual-world task. In the priming task bilinguals displayed inhibition for higher-frequency targets and a trend towards facilitation in lower-frequency targets. Increased L2 proficiency led to decreased lexical competition for bilinguals in the visual-world task. Despite predictions of the bilingual advantage literature, domain-general inhibitory control did not appear to have any impact on lexical competition, and there was no bilingual advantage in domain-general inhibitory control. Below I discus each task in further detail as well as the implications for the findings relating to domain-general inhibitory control and L2 proficiency.

Visual world

The relatedness manipulation in the visual world-task was successful in three of the four outcomes — RT, AUC, and MD — across all speakers, therefore lexical competition was induced in this task both in monolinguals and late bilinguals who were processing their L2. Within the related trials, the manipulation of relative target-competitor frequency was evident in the RT for both groups of speakers and in the mouse trajectories of bilinguals. English proficiency in the bilingual group also impacted the degree of lexical competition; those with higher proficiency moved towards the competitor to a lesser degree on the way to the target.

Generally, the entrenchment account predictions were most compatible with the results. Frequency did not affect monolingual mouse trajectories during related trials, but bilinguals did tend to veer more towards higher-frequency competitors. While a speaker by frequency interaction was not demonstrated in models with all speakers, differences between the speaker groups appeared in separate models for monolinguals and bilinguals. One of the seminal findings of the entrenchment account is larger frequency effects in monolinguals with lower entrenchment (Kuperman & Van Dyke, 2011). Due to the fact that the late bilinguals in the current study had less overall exposure to English compared to the monolinguals (i.e., lower subjective English frequency), the entrenchment account predicted that bilinguals would display larger frequency effects. This pattern occurred in the mouse-trajectory patterns (i.e., a more direct measure of lexical competition than RTs), where bilinguals showed an effect while monolinguals did not.

There are a few possible reasons why the speaker by frequency interaction did not appear as strongly as expected. One could be the relatively high proficiency in the bilingual group, which was at a level of Advanced Mid on the ACTFL scale or greater. The entrenchment account predicts that bilinguals with lower proficiency may show an even greater effect of frequency. Another reason may be that the predictions for monolinguals were based on eye-tracking (e.g., Dahan & Gaskell, 2007). While mouse-tracking can capture a more continuous cognitive process than eye saccades, perhaps eye movements are further upstream or more automatic, and therefore more sensitive, than mouse movements. The third reason is a that there may have been some related target-competitor pairs whereby the frequency approximated in the SUBTLEX-US corpus did not reflect subjective frequency among the participants. For instance, the corpus indicated that *butter* was lower-frequency than *bubble*. While that is true in the corpus itself (i.e., in the subtitles of television and films), it may be that *butter* is encountered more often in daily life (e.g., around mealtimes), therefore not captured in this corpus. Although the by-item random intercepts and slopes was able to incorporate this sort of item variation, a set of stimuli without this variation at all would provide a stronger test of frequency.

A large portion of the entrenchment account discusses how proficiency can be viewed as a proxy for L2 entrenchment. For instance, in Diependaele, Lemhöfer, & Brysbaert (2013), performance on a visual progressive demasking task was partially accounted for by LexTALE scores, more so than it was by the participants' native language. Bilingual participants in the current study with greater L2 proficiency were better able to manage the lexical competition generated by phonological overlap. It should be noted that the proficiency pattern is not incompatible with the bilingual advantage literature. Nonetheless, proficiency is never directly addressed in the bilingual advantage literature as most bilinguals are early or even simultaneous language learners (i.e., learn L1 and L2 concurrently), and proficiency may be near ceiling and less varied in these types of bilinguals.

The current study is novel in that it attempted to address the bilingual advantage with late L2 learners. The material selection, therefore, was limited to the top 5,000 most frequent English words to ensure the stimuli were known by the participants. In that sense, this criterion was successful since 98.52% of the words were rated as known or well-known by the bilinguals. At the same time, this criterion led to a reduced frequency range in the present study compared to previous research.

The average log frequency according to the SUBTLEX-US corpus in the higher-frequency targets was 75.54 instances per million (pm) and 15.26 pm in the lower-frequency targets, with a range of 8.02 pm to 434.60 pm. For comparison, the L2 English stimuli in Duyck, Vanderelst, Desmet, and Hartsuiker (2008) in the lower-frequency condition had an average frequency of 2.18 pm and 9.21 pm in the higher-frequency condition, according to the CELEX database (R. H. Baayen, Piepenbrock, & van Rijn, 1993). This visual-world study demonstrated large L2 frequency effects in Dutch-English bilinguals in addition to smaller L1 frequency effects in the same participants.

Since the familiarity task revealed that the late learner population was highly familiar with the top 5,000 most frequent English words and there were effects of frequency even in this restricted range, follow-up studies could include a wider frequency range for a similar population. Once a larger material pool is available, not only will this allow for greater frequency differences to appear, but other lexical characteristics can be explored. For instance, frequency and neighborhood size could be more easily orthogonally varied with a wider selection of lexical items.

While not all interactions were as considerable as expected, several outcomes of the study contribute to the existing body of data on bilingual lexical access. The late and highly proficient bilinguals were distracted by phonological competitors in L2 and to a similar degree as monolinguals. However, bilinguals were not a monolithic group. They were slightly more sensitive to the relative frequency of target and competitor, and those with lower proficiency were less efficient in handling lexical competition overall.

Phonological priming

In the phonological priming task, there was no evidence of an overall priming effect (i.e., related trials slower than unrelated trials) across speakers. However, once speaker group was added as a fixed effect, some predicted patterns began to emerge. Bilinguals displayed a pattern of facilitation in the lower-frequency trials and inhibition in the higher-frequency trials. Although not strong, these findings replicate those in Cook and Gor (2015) and Gor and Cook (2018). The overall high frequency range in the current priming stimuli may have again reduced the effects in bilinguals, as in the visual-world task. Targets ranged in frequency from 2.00 pm to 245.27 pm, with an average of 17.40 pm. For comparison, the average Russian target lemma frequency for the same inter-stimulus interval (350 ms) in Cook and Gor (2015) was 5.98 pm with a range of 4.48 pm to 12.18 pm. In the same study, primes that were less well-known produced facilitation. In the current study, only trials with known primes and targets were included in the analysis (although note that being able to identify translations as correct or incorrect would add strength to the familiarity ratings that participants provided). Thus, the flip from inhibition to facilitation is even more striking given that all primes and targets were reported to be known.

One tentative conclusion is that familiarity with, and therefore entrenchment of, a nonnative lexical representation is best thought of as a continuum, and proficiency and frequency interact within individuals to influence the degree of entrenchment. Lower-frequency words tend to also be less familiar, and lower proficiency bilinguals will have encountered these lower-frequency words even less frequently. Along the continuum, as entrenchment increases so does the strength of a representation as a competitor.

Broersma (2012) was another study with a mix of nonnative inhibition and facilitation during priming. The purpose of the experiment was different (i.e., to investigate confusable phonemes between primes and targets such as in *flesh-flash*), the materials were lower in frequency (i.e., average frequency of 5.67 pm), and the participants were native English or Dutch-English bilinguals. Broersma (2012) found no differences between related (e.g., *flesh-flash*) and unrelated conditions on average for the bilinguals. Because the words in the related condition contained confusable phonemes, it is possible that the primes were not well entrenched and therefore not able to exert full inhibition on the targets. Paired with the lower frequency, the author suggests that the mix of inhibition and facilitation from the related items caused a null effect in the analysis. The phonolexical quality of a representation therefore appears to also determine placement along the familiarity continuum.

In both tasks, it appears that bilinguals were more sensitive to frequency than monolinguals. The high frequency range may have led the monolinguals to exhibit faint frequency effects, in conformity with the entrenchment account. On the other hand, the lexical inhibition scores between the visual-world and priming tasks were not correlated even when separated by speaker type, highlighting the difference in the type of lexical competition between the two tasks. Competitor inhibition was observed in bilinguals in the visual-world task where the competitor was visually present when hearing the target. Bilinguals exhibited a range of facilitation to inhibition in the priming task when the target was one of many 'virtual' competitors during prime recognition. Conversely, monolingual displayed the most robust evidence of lexical inhibition in the visual-world task, where the presence of a direct cohort competitor during target recognition induced lexical competition despite the high frequency stimuli.

One of the more surprising results in the phonological priming task was the lack of a priming effect in monolinguals. There were trends for inhibition in the lower-frequency trials and facilitation in the higher frequency trials. One explanation is that the stimuli were all too high frequency for this group of participants. This finding may be an example of the attenuated frequency effects predicted by the entrenchment account taken to an extreme. Higher-frequency words presented to individuals with entrenched lexicons lead to very fast recognition to such a degree that competitors are barely activated. Segui and Grainger (1990) also did not observe priming when primes were higher in frequency than targets, albeit in the visual modality, suggesting that only competitors with more activation than the target are inhibited. There could be a frequency level at which frequency does not impact competition if both target and competitor are considered equally frequent. Under the entrenchment account, this tipping point is lower for individuals with more entrenched lexicons.

In hindsight, it may have been better to have included an expanded lower-frequency range, as that is where monolinguals tended to display inhibition of a target after a phonologically related prime. A follow-up study could provide monolinguals with a different set of words, encompassing a wider frequency range, especially on the lower end. However, the disadvantage in this case is that there could not be a direct comparison between monolinguals and bilinguals.

Another possibility is that the task order led to an element of strategy. The priming task always followed the visual-world task, with the Stop Signal task intervening. It is possible that the participants became aware of the phonological overlap as the focus of the “language tasks,” as they were characterized to participants, after encountering related trials in the visual world task. However, of the 50 visual-world trials, 16 (30%) were phonologically related, and this ratio was even lower in the priming task (16% of trials contained phonological overlap). These ratios are common in previous studies to mask the phonological overlap manipulation, but perhaps two tasks together make it more apparent.

In the priming task, bilinguals displayed inhibition in higher-frequency trials, indicating they did experience lexical competition. Monolinguals did not display many signs of lexical competition in this task. A tentative conclusion and extension of the entrenchment account is that strength of a lexical representation can lead to not only facilitation and inhibition but also to very little inhibition for well-entrenched lexicons. Frequency and phonolexical quality affect familiarity, and L2 words that are more familiar are stronger competitors.

Individual differences

The predictions for the individual difference measures at the group level were that bilinguals would exhibit better domain-general inhibitory control, or at least be equivalent to monolinguals (e.g., Bialystok, Martin, & Viswanathan, 2005), and that both groups would be similar in fine-motor skills with a computer mouse. Domain-general inhibitory control was measured via three reliable tasks which were combined into one composite score to capture the inhibitory control component common across the three tasks (i.e., to partially remove task-specific skills not related to inhibitory control). The pursuit rotor task was utilized to assess fine-motor skills with a computer mouse.

Monolinguals in the current study displayed better domain-general inhibitory control and better fine-motor skills with a computer mouse. It appears that using language (i.e., the L2 for bilinguals) in any of the individual inhibitory control tasks was not driving the differences, as bilingual English proficiency did not correlate with any of the individual tasks nor the composite, and monolinguals still outperformed bilinguals in the Antisaccade and Stop Signal task.

This outcome initially leads to a conclusion of a monolingual advantage in domain-general inhibitory control for the current group of participants. However, the results of the pursuit rotor task temper that conclusion. Monolinguals were not only better on both outcomes of the pursuit rotor task (i.e., pixel deviation from the red dot and time on target), but they were also faster during the control or neutral trials of the Stop Signal and Stroop tasks, respectively. This pattern differs from the bilingual advantage, where children were better at Stroop congruent and incongruent trials but equivalent to monolinguals during control trials (Bialystok, 2010) Therefore, the monolingual advantage observed for the inhibitory control composite may in part be due to better performance overall on a computer.

One speculative explanation is length and intensity of exposure to a computer in daily life. The monolinguals were native English speakers and were born and matured in the United States. Bilinguals were more heterogenous as a group and, although information about computer experience was not incorporated in the language history questionnaire, it is possible that bilinguals had a wider variety of computer experience over their lifetime than the monolinguals. In most other aspects (e.g., education level, age), the groups were relatively similar.

There are indications in the bilingual advantage literature of a link between lexical and domain-general inhibitory control (e.g., Blumenfeld & Marian, 2007), which led to the prediction in the current study that domain-general inhibitory control is a mechanism by which lexical competition is managed. Explicitly discussed only in the bilingual advantage realm, the prediction is consistent with the entrenchment account as well — those with better inhibitory control, regardless of speaker type, should show smaller effects of lexical competition in both tasks (e.g., smaller differences between related and unrelated trials in the priming task). It was also predicted to interact with speaker group in a cumulative fashion, such that those individuals with better domain-general inhibitory control in the group that was more efficient at managing lexical competition were expected to show the smallest signs of competition.

Given the lack of strong group differences in the linguistic tasks gauging the degree of lexical competition throughout the study, there was low probability of inhibitory control showing an effect. Monolinguals displayed better domain-general inhibitory control, thus they were the group most likely to be able to bring it to bear during difficult lexical retrieval situations (i.e., strong lexical competition). However, the items in both tasks were drawn from a fairly high frequency range, thus strong competition modulated by frequency was not seen in either task, nor was competition seen at all in priming in the monolingual group.

If monolinguals were to utilize domain-general inhibitory control, it would be in situations of many and/or strong lexical competitors. For instance, the lexical and domain-general inhibitory control link may appear monolinguals with smaller vocabularies (i.e., low entrenchment) or in the lower- to mid-frequency range. Following the idea that higher-frequency ranges did not induce competition for individuals with well-entrenched lexicons in the priming task, evidence of domain-general inhibitory control at play in monolinguals would most likely occur in lower-frequency representations or in individuals will less-entrenched lexicons overall.

Lexical competition was observed in bilinguals, although a relationship between it and domain-general inhibitory control were less likely to appear in this group due to their lower inhibitory control abilities (i.e., it was a less powerful resource to utilize during lexical access). Moreover, since lower-frequency items were excluded from models including the inhibitory control composite, there were fewer items available to analyze with regard to the relationship between lexical and domain-general inhibitory control in the priming task.

Putting aside the specific parameters and outcomes of the present study, the bilingual advantage literature predicted not only an inhibitory control advantage in bilinguals but also an association between lexical and domain-general inhibitory control. The bilingual participants in this study had high L2 proficiency, but unlike previous studies, were late learners instead of early or simultaneous bilinguals. This late learner group displayed no evidence of a bilingual advantage in inhibitory control nor its use in lexical competition. There have recently been calls to further investigate bilinguals not as one monolithic group but to understand the role of individual experiences in language learning and the bilingual advantage (Baum & Titone, 2014), and the results of the present study add age at the onset of learning as a separate factor to explore. Thus, the present study has not elucidated an association between lexical and domain-general inhibition in either speaker group.

Proficiency

English proficiency, measured by the online LexTALE (Lemhöfer & Broersma, 2012), was examined only in bilingual participants. Scores on this task impacted RTs in the phonological priming task, although only as a main effect. Participants with higher L2 proficiency were faster. Proficiency did not influence the degree of the priming effect, again showing that lexical competition in this task was more subtle than in the visual-world task. LexTALE score affected the two measures of mouse trajectory deviation (i.e., area under the curve and maximum deviation) during related trials of the visual-world task; those with higher English proficiency veered less towards the competitor on the way to the target. When there was a direct competitor present, lexical competition was better managed by those with better proficiency.

The entrenchment account references proficiency specifically as a method for measuring degree of entrenchment – those with more entrenched lexical representations have higher proficiency. The bilingual advantage does not often discuss proficiency, but most studies use early or simultaneous bilinguals (i.e., individuals for whom proficiency is near ceiling). Perhaps another reason domain-general inhibitory control did not impact lexical competition in bilinguals is because proficiency was still at play. That is, it is possible that these two accounts depict different ends of the language-learning spectrum. Entrenchment captures early learning while the L2 lexicon is still being entrenched, during which the goal is to establish stronger L2 representations. As representations become more entrenched, L2 and L1 lexical items become equally strong competitors, which may be when domain-general inhibitory control becomes a useful tool to manage competition. Exactly where this point is along the language-learning journey depends highly on the circumstances; for instance, do L1 and L2 contain confusable phonemes? This idea is consistent with recent reviews of the bilingual advantage literature stating that bilinguals are not a homogeneous group when it comes to use of inhibitory control (e.g., Baum & Titone, 2014). Additionally, including bilinguals with various L1s potentially created a bilingual group that was less homogeneous in their L2 experiences as well as in early exposure to computers and computer games. Bilingual participants had different L2 learning profiles, e.g., degrees of intensity and lengths of time learning L2 while not in the United States, which may have mitigated the role of L2 proficiency and inhibitory control in lexical competition. Notwithstanding, proficiency still played a role in the current study.

The degree to which monolinguals use domain-general inhibitory control during lexical access is still an empirical question. Adding more nuance to the bilingual advantage narrative, it may not be that monolinguals do not use domain-general inhibitory control to manage lexical competition, just not to the same degree as bilinguals. As previously discussed, there may still be situations where monolinguals could utilize domain-general inhibitory control (e.g., lower-frequency words), but these situations do not occur as often as for bilinguals. Of course, this interpretation assumes that all bilinguals reach a stage of L1 and L2 competition due to strong L2 entrenchment, which may not be the case. These conjectures are speculative and further research is necessary for a more comprehensive understanding of lexical competition.

Conclusion

Across both tasks, bilinguals were sensitive to frequency, and monolinguals were less sensitive or did not show any frequency effects. The entrenchment account predicted that bilinguals would be more susceptible to lexical frequency since bilinguals would have a lower subjective frequency for all words, but especially so for lower-frequency words (i.e., similar to monolinguals with smaller vocabularies in Kuperman & Van Dyke, 2011). Entrenchment was weak enough in the late bilinguals to cause facilitation in the lower-frequency items in priming, suggesting that the primes were not inhibiting the target competitors but activating the sublexical components in common with the target.

In the visual-world task where the competitor was visually salient and the task necessitated activation of semantic information, both groups displayed signs of lexical competition, and the higher-frequency competitors affected the bilinguals to a greater degree than the monolinguals. Despite the relatively high frequency of the items in all tasks for monolinguals, lexical competition was still observed in the visual-world task. These monolingual findings are consistent with the types of lexical competition the two tasks were designed to measure. That is, the visual-world task imposed lexical competition that had to be managed while the phonological priming task provoked target inhibition via normal lexical access processes.

The bilinguals in the current study were late learners, as opposed to early or simultaneous bilinguals in most bilingual advantage studies (e.g., Bialystok, Craik, & Luk, 2008). Domain-general inhibitory control did not exhibit strong influence in the bilinguals in the current study; conversely, variation in lexical competition management was somewhat accounted for by proficiency and lexical characteristics. The relationship between domain-general and lexical inhibitory control may still exist in bilinguals but may require time to develop, which is why it was not observed in the current study. Accordingly, the results with late bilinguals fit with the bilingual advantage idea that increased domain-general inhibitory control is related to experience with increased lexical competition. The late bilinguals in the current study lacked this experience. Future studies may directly compare early bilinguals, late bilinguals, and monolinguals with stimuli encompassing a greater frequency range in order to expand on the current findings.

Appendices

Appendix A: Pilot Studies

While the materials for the current study were being selected and normed, materials from a preceding but very similar study (Lancaster, 2017) were used to pilot the timing of two proposed mouse-tracking trial types: visual-world and priming probe. Specifically, we examined the timing necessary for each given that we could not find any extant literature that used a priming probe trial following a visual-world trial with mouse tracking.

The goal of the priming probe trials was to assess lexical competition during the visual world trials, as were conducted in Blumenfeld & Marian (2011) and again in Blumenfeld, Schroeder, Bobb, Freeman, & Marian (2016). In these pilot studies, each visual-world trial was followed by a priming probe trial, which displayed two asterisks in place of the two previously displayed pictures. The task was to click on the grey asterisk, which sometimes replaced the target and sometimes replaced the competitor. The logic behind these trials is that negative location priming causes participants to be slower when responding to locations in which pictures were previously inhibited (i.e., competitors) and faster when responding to locations in which pictures were previously selected (i.e., targets).

In Blumenfeld & Marian (2011), monolinguals were faster when responding to probes (i.e., grey asterisks) in previous target positions compared to probes in previous control positions (i.e., a filler item in a four-picture visual-world trial). Competitor probes were slower than control, indicating that competitors were inhibited during visual-world trials and that targets were selected (i.e., not inhibited) during previous trials. The pilot included only two-picture visual-world trials, which is common in the mouse tracking literature (e.g., Incera & McLennan, 2016; Spivey et al., 2005) because it increases sensitivity due to a greater area for mouse trajectories to deviate. Since there were no filler pictures (i.e., only target and competitor, related and unrelated), there were no control priming probe trials.

Given the results of Blumenfeld & Marian (2011), I predicted that target probe reaction time (RT) would be greater than competitor probe trials RT. Although the software captured all four mouse-tracking outcomes for the priming probe trials, only RT was analyzed, as it was the outcome used in Blumenfeld et al. (2016). A novel extension of the priming probe, had the pilot trials been successful, would be to compare RT of not only previous target and competitor locations, but to compare RT of previously related and unrelated competitor locations.

While two-picture, visual-world trials demonstrating the effect of phonological relatedness with mouse tracking are common (e.g., Spivey et al., 2005), the main impetus of the pilot was the unknown effect of interleaving priming probe trials with the visual-world trials in this methodology. Thus one of goals of the pilot studies was to determine if including priming probe trials affected the robust phonological relatedness effect during visual-world trials. Another goal was to determine the amount of time necessary between when the pictures first appear (either asterisks for the priming-probe trials or objects for the visual-world trials) and when participants can move the mouse cursor. A 500 ms asynchrony is usually imposed during visual-world, mouse-tracking trials to ensure participants do not wait until the entire word is heard before moving the mouse.

The software utilized, MouseTracker (Freeman & Ambady, 2010), imposes constraints that affect presentation timing and also necessitated the pilot studies. The first constraint is that it does not allow differential timing within a list. That is, all trials, visual-world and priming, were required to have the same amount of asynchrony between the appearance of pictures and the ability to respond. Secondly, it is always necessary to have participants click on a button with the world *Start* to begin a trial, regardless of trial type. While the inter-trial interval was programmed to be 250 ms, the amount of time to click the *Start* button varies from participant to participant and trial to trial. Thus, it is possible that the inter-trial interval is greater than 250 ms. The timing of the inter-trial interval is important because lexical inhibition, especially that due to phonological relatedness, decays quickly, with no evidence of lexical inhibition found after 500 ms (Goldinger, Luce, & Pisoni, 1989).

Methods

Two pilot studies were conducted. The first did not include any time between when the pictures appeared and when participants could move the mouse cursor to select a target (i.e., the 0 ms pilot). This timing would favor the priming probe trials due to a shorter delay between the previous visual-world trial and the priming-probe trial, preventing activation and/or inhibition decay. Twenty-five native speakers of English completed this version of the pilot mouse-tracking task. The second pilot included a 500 ms asynchrony (i.e., the 500 ms pilot), favoring the visual-world trials by allowing participants time to recognize the object depicted before the target is heard. Fourteen different native speakers of English completed this version of the pilot mouse-tracking task.

The materials for both pilot studies were identical. One of the lists from Lancaster (2017) was modified to include pilot-probe tasks. Participants first read instructions about how to respond to each trial type (i.e., click on the grey asterisk or the picture that corresponds to the word you hear) then completed 10 practice trials consisting of five visual-world and five priming-probe trials. The location on the screen (i.e., right or left top corner) of visual-world target and subsequent priming-probe target was balanced throughout the list. The list contained 20 phonologically related trials with similar characteristics as the study presented in the main text (e.g., *muffin – monkey*) and 39 phonologically unrelated trials (e.g., *lantern – slipper*), which were a combination of filler and unrelated trials from Lancaster (2017). Thus there were also 59 priming-probe trials. Then entire list took approximately five minutes to complete. Table 22 lists all the visual-world trial materials grouped by trial type.

Table 22: Pilot visual-world materials

Related Trials		Unrelated Trials		Filler Trials (unrelated)	
Target	Competitor	Target	Competitor	Target	Competitor
fries	frog	backpack	genie	bagpipes	wheelbarrow
corn	cork	grasshopper	saxophone	tugboat	bookshelf
crab	crown	camel	trumpet	tassel	wishbone
mask	match	clam	wrench	ponytail	tweezers
monkey	muffin	clock	boat	eggplant	sandbox
note	nose	lantern	slipper	lollipop	clarinet
parachute	pajama	lifeboat	dresser	astronaut	fingerprint
pillow	pitcher	balcony	microphone	lizard	recorder
platter	pliers	panda	ashtray	egg	grave
bat	badge	purse	nest	angel	fire
shell	chef	ram	fence	gazebo	headlight
skull	scale	rainbow	scissors	leprechaun	chandelier
scorpion	skateboard	raincoat	parrot	doorknob	swimsuit
spatula	spaghetti	road	dress	moose	triangle
star	stairs	skunk	axe	pirate	strawberry
toothpick	tuba	square	pen	witch	plant
blimp	blinds	toothbrush	boxer	gate	queen
bowl	bow	walnut	beetle	eye	gun
diamond	diaper	desk	bread	knife	tree
flashlight	flower	dragon	medal		

Results

For all analyses, trials with incorrect responses were excluded as were trials greater or less than three standard deviations per participant. Linear multilevel models were also conducted with the most recent version of the lme4 package (Bates et al., 2016), and all models were run as forced-entry models for fixed effects and included cross-classified subject and item random intercepts. Random slopes were not tested due to the small sample size, which may have prevented model convergence. Since only *t*-values are provided in the lme4 output and the methods for computing a *p*-value remain controversial, *t*-values with an absolute value greater than 2.0 are considered significant and greater than 1.65 are considered marginally significant (Gelman & Hill, 2007). All models utilized treatment coding for categorical variables (i.e., related vs. unrelated visual-world trials or target vs. competitor pilot-probe trials).

Table 23 provides the average response times per probe type in each pilot study. The first pilot, with 0 ms asynchrony between picture appearance and the ability to move the mouse cursor, did not demonstrate the predicted results for either trial type. Exclusions due to incorrect responses and trimming resulted in 1.08% of data being excluded for the priming probe analyses and 0.90% being excluded for the phonological relatedness analyses. During the priming-probe trials that followed non-filler visual-world trials, probes that replaced a previous visual-world target, or target probes, were not more quickly selected than probes that replaced a previous visual world competitor, or competitor probes (b = -10.24, SE = 13.29, t = -0.77; competitor as reference). The second, and novel, portion of this analysis predicted that when the competitor probes were divided into those that followed related and unrelated trials, related competitor probes would be slower. This prediction was borne out (b = -34.17, SE = 15.90, t = -2.15; related competitor as reference). Thus the replication of Blumenfeld and Marian (2011) failed in this methodology (i.e., competitor probes did not have slower RTs than target probes), but comparing probe types (i.e., related and unrelated) had promise as an avenue for demonstrating inhibition in a previous visual-world trial.

Table 23: Average (SD) reaction times per condition for pilot priming-probe trials

	Competitor All	Target	Competitor Related	Competitor Unrelated
Pilot 1 (0 ms)	820.90 (165.67)	811.12 (152.63)	838.22 (174.05)	803.85 (155.38)
Pilot 2 (500 ms)	844.33 (294.35)	819.41 (266.96)	859.11 (288.39)	829.56 (300.29)

With regard to the visual-world trials, there was very little evidence that phonological overlap (i.e., the related trials) had an effect during the first pilot. Table 24 details the average outcomes in each pilot study by trial type (i.e., related or unrelated). Related trials serve as the reference for all reported models. Related visual-world trials were not significantly different than unrelated trials in terms of reaction time (b = 1.47, SE = 34.08, t = 0.04), AUC (b = -0.1912, SE = 0.1288, t = -1.48), MD (b = -0.0817, SE = 0.0561, t = -1.46), nor flips (b = -0.0031, SE = 0.2783, t = -0.01). Differences in MD and AUC between these two trial types, however, were at least trending in the correct direction (i.e., larger values, or more distraction towards the competitor, in related trials), although these differences were not significant with 25 participants.

Table 24: Average (SD) visual-world outcomes by trial type

Outcome	Condition	Pilot 1 - 0 ms	Pilot 2 - 500 ms
RT	related	1121.38 (268.82)	1362.36 (320.92)
	unrelated	1124.21 (267.95)	1348.20 (378.93)
AUC	related	0.9340 (1.2112)	0.2902 (0.5075)
	unrelated	0.7436 (1.0810)	0.2012 (0.4100)
MD	related	0.4903 (0.4841)	0.2902 (0.5075)
	unrelated	0.4090 (.4657)	0.2012 (0.4100)
Flips	related	5.96 (3.45)	7.01 (3.73)
	unrelated	5.96 (3.15)	7.07 (3.84)

The second pilot, with 500 ms asynchrony between picture appearance and the ability to move the mouse cursor, additionally did not demonstrate the predicted results for either trial type. Exclusions due to incorrect responses and trimming resulted in 2.06% of data being excluded for the priming probe analyses and 1.25% being excluded for the phonological relatedness analyses. During the priming-probe trials that followed non-filler trials, target probes were not more quickly selected than competitor probes (b =

-24.52, *SE* = 29.26, *t* = -0.84; competitor as reference). Unlike the first pilot, responses to related competitors were not significantly slower than responses to the unrelated competitors (*b* = -33.79, *SE* = 35.11, *t* = -0.96; related competitor as reference).

As with the first pilot study, there was little evidence that phonological overlap (i.e., the related trials) had an effect during the first pilot. Related trials serve as the reference for all reported models. Related visual-world trials were not significantly different than unrelated trials in terms of reaction time (*b* = -14.25, *SE* = 31.04, *t* = -0.46) nor flips (*b* = -0.06, *SE* = 0.35, *t* = 0.16). AUC, however, was one of the outcomes to significantly demonstrate greater competitor influence in related compared to unrelated trials (*b* = -0.2235, *SE* = 0.0983, *t* = -2.27), as was MD, albeit marginal (*b* = -0.0877, *SE* = 0.0498, *t* = -1.76). Both outcomes are in the predicted direction, that is, related competitors caused mouse trajectories to veer more towards the competitor than the unrelated competitors.

Discussion

In first pilot, not only did the phonological effects not appear, but the main effect of priming probe, target compared to competitor, also was not evident. Thus the 0 ms asynchrony between picture appearance and the ability to move the mouse cursor was not deemed viable. That is, not only were the new, priming-probe trials not demonstrating predicted effects, but the timing was eliminating the robust effects of phonological overlap.

The second pilot did not bode well for the priming-probe trials. While the 500 ms asynchrony did bolster the visual-world trials such that half of the outcomes did show the predicted effects, the effects were not as robust as when the priming-probe trials were not present (Lancaster, 2017). Moreover, the priming-probe trials were not serving their purpose of differentiating inhibition in previous competitors and activation in previous targets with this asynchrony. Therefore, it was decided to remove the priming-probe trials from the experimental design. Although they were meant to investigate the degree of lexical inhibition after a trial, a multifaceted approach was still possible given that the phonological priming task was able to examine a similar situation.

Appendix B: Language history questionnaire

The following is a paper version of the online language history questionnaire, with the exception of the consent form, which was the first part of the questionnaire.

1. Gender
    a. Male
    b. Female
    c. Other
2. I am
    a. Right-handed
    b. Left-handed
3. Age: __________ years old
4. Year in college:
    a. Freshman
    b. Sophomore
    c. Junior
    d. Senior
    e. Graduate student
    f. Not a student
5. My first language, meaning the language I've been exposed to and spoken since birth, is (choose all that apply):
    a. English
    b. Spanish
    c. Other: ___________________
6. Estimate, in terms of percentages, how often you use your native language and other languages per day (in all daily activities). Total should equal 100%
    a. Native language (slider 0-100%)
    b. Second language, please specify (slider 0-100%)
    c. Third language, please specify (slider 0-100%)
    d. Fourth language, please specify (slider 0-100%)
7. The second language I learned (if any) was: ___________________________
8. How old were you (years) when you started learning your second language (listed above)?
    a. ___________
9. If you have lived in a country where most people speak your second language (listed above), please enter how long you lived there (in months) and the age at which you arrived (in years)
    a. Length of stay (months) ___________
    b. Age (years) ___________

10. If you have had any formal instruction in your second language (listed above), please mark in which context (e.g., in college) and for how long (e.g., 2 years) in the chart below.

	1 year	2 years	3 years	4 years	5+ years
In elementary school (grades 1-3/4)					
In middle/high school (grades 5-11)					
In college					
With a private tutor or attended language courses					

11. Please rate your proficiency in your second language (listed above) in each of the following domains on a scale of 1 (minimal) to 10 (like a native speaker).

	1 (minimal)	2	3	4	5	6	7	8	9	10 (native-like)
Speaking										
Pronunciation										
Listening										
Reading										
Writing										
Grammar										

12. The third language I learned (if any) was: ____________________________
13. How old were you (years) when you started learning your third language (listed above)?
    a. ___________
14. If you have lived in a country where most people speak your third language (listed above), please enter how long you lived there (in months) and the age at which you arrived (in years)
    a. Length of stay (months) ___________
    b. Age (years) ___________

15. If you have had any formal instruction in your third language (listed above), please mark in which context (e.g., in college) and for how long (e.g., 2 years) in the chart below.

	1 year	2 years	3 years	4 years	5+ years
In elementary school (grades 1-3/4)					
In middle/high school (grades 5-11)					
In college					
With a private tutor or attended language courses					

16. Please rate your proficiency in your third language (listed in Question 6) in each of the following domains on a scale of 1 (minimal) to 10 (like a native speaker).

	1 (minimal)	2	3	4	5	6	7	8	9	10 (native-like)
Speaking										
Pronunciation										
Listening										
Reading										
Writing										
Grammar										

17. Please list any other languages you know and how well you know them overall on a scale from 1 (minimal) to 10 (like a native speaker) (e.g., Portuguese – 3, German - 1)

Appendix C: Procedure

The following is the full procedure for the in-person experiment. It was used to train research assistants and was present when each participant completed the tasks.

Before the Subject Arrives

- Start and log into computers (in Room A or B)
- Open PsychoPy by navigating on the PC to **C:/Desktop/Alia/diss_exp** and click on **PsychoPy2.exe**
  - Sometimes in room B it looks like psychopy is open but you cannot see the window. To get the window back, press shift and right click on the psychopy icon on the taskbar (bar on bottom of screen). Then release the shift button and either more the mouse or press the right arrow key (or some combination of both). A tiny window will appear on the left side of the screen, which you can then move and maximize.
- Make sure the PC has headphones plugged in and the black mousepad under the mouse
- Gather testing materials
  - For everyone:
    - Participant tracking sheet
    - Protocol
  - Payment log for paid participants

Outline of study task order

	task	lists	computer	time	speaker group to complete task
1	mouse tracking	mt_a, mt_b, mt_c, mt_d (refer to participant tracking sheet)	PC	5 minutes	both
2	stop-it	none	PC	15 minutes	both
	BREAK (optional)	------------------------------	----------	---------	----------
3	priming	pp_a, pp_b (refer to participant tracking sheet)	PC	10-15 minutes	both
4	pursuit rotor	none	PC	2 minutes	both
	BREAK (optional)	------------------------------	----------	---------	----------
5	stroop	none	PC	8 minutes	both
6	antisaccade	none	MAC	15 minutes	both
7	familiarity	none	PC	25 minutes	L2 only

After the Subject Leaves

1. Put completed payment paperwork in file cabinet
2. Replace protocol and subject tracking sheets

When the Subject arrives

Greet the participant:

Hi, are you (participant's name)? Great, my name is (your name) and I will be working with you today.

Before we begin, I want to remind you to <u>*turn off your cell phone*</u> *and if you are chewing gum, please throw it away. I also want to say that your help and attention is very important to our project, so I ask you to devote your full attention to what we will be doing today, and that you should try and do your best.*

I'm going to have you do a variety of tasks today. You will do some language tasks and cognitive tasks. We'll switch between the tasks so you don't get too bored with any one task. There are also 2 points at which you can take a 5 minute break if you want. Some of the tasks are fairly easy and some of them are rather difficult. Please follow <u>*all*</u> *of the instructions I give you. Do you have any questions?*

Please adjust the screen and keyboard so they are centered in front of you.

Notes

- Always sit by the subjects so you can ask and make sure they understand the instructions. Let the subjects read the instructions for themselves. For the antisaccade task, you will need to advance to the next screen, but the other tasks are self-paced. Stay in the room so that the participant actually performs each task and to make setting up each task easier.
- Always make sure they know which buttons to use (L & R) and which button stands for which response. For all tasks, they should use both hands, only responding with **index fingers**. Always watch them during the practice to make sure they are doing the task correctly. If they do not understand, you can redo the practice by quitting the task and starting over by hitting COMMAND + PERIOD (antisaccade) or ALT+F12 (mouse tracking) or ESC (all other tasks).
- Most tasks require you to enter the participant number then press ENTER to begin the task. Some programs won't recognize the ENTER key by the number pad – best to stick with the numbers and ENTER key by the letter keys.
- If a subject is COLORBLIND, then have them try to do the color tasks (stroop). If it is clear that the subject is having trouble distinguishing colors, just skip the task and move on to the next one
- At the end of each task, when you quit you may be asked if you want to save changes. ALWAYS CLICK NO.
- When running pursuit rotor or STOP-IT, a window may pop up asking for permission or something else. Press "yes," or the equivalent.

Mouse Tracking

- Have participant sit in front of PC. Make sure to use the black mousepad and position the mouse and mousepad in the center of the screen.
- Have participant put on headphones and adjust sound to a comfortable level. They will not able to adjust the sound once the task begins.
- During the task, make sure the participant is clicking on the correct picture (if you can hear through the headphones). If they are not clicking on the correct picture or not responding, tap them on the shoulder and remind them to answer as quickly and as accurately as possible.

Open one of the lists (refer to the participant tracking sheet):

1. Go to the folder **Desktop\Alia\diss_exp**
2. Click on **MT runner.exe** (the icon is a lightening bold) to open the mouse tracking software. Note that it is called just **runner.exe** in the annex.
3. A window will pop up asking you which list to open.
    a. Choose either mt_a, mt_b, mt_c, mt_d, depending on what is listed in mouse tracking task on the participant tracking sheet.
4. Enter the subject number and push enter

Provide instructions:
In this task, you will see 2 pictures on the screen and hear a word corresponding to one of the pictures. Use the mouse to click on the picture corresponding to the word you hear. If you click on the wrong picture, a red "x" will appear. If you do not click on any picture, the words "no response" will flash on the screen. Please try to answer as quickly and as accurately as possible.

You'll see some instructions and then get to do some practice before the real trials begin.

STOP-IT

- Have the participant sit in front of the PC.
- Headphones will remain on.
- Make sure the volume is set at 50
- Left key = Z   Right key = ? /
- They should use both hands (index fingers) for pressing the left and right arrow keys. Press the SPACE BAR to start. There will be a short practice. At the end, there will be a screen presenting their average RT and % correct.

Open the task:

1. Go to the folder **Desktop\Alia\diss_exp**
2. Click on the **STOP-IT** shortcut to start the task.
3. Enter the subject number and push enter

Provide instructions:
In this task, you will see either a square or a circle. Your task is to push the left key (Z key) if you see a square, and the right key (?/ key) if it is a circle. Before each shape,

you'll see a small plus sign, which is just a fixation point. The shapes will disappear quickly, so try to answer as quickly and accurately as possible.

Sometimes you will hear a beep after you see a shape. When that happens, try to not respond (don't press anything), and just wait for the shape to disappear from the screen. Sometimes you will not be able to stop, and that's ok. The program was written so that the amount of time before you hear the beep changes, making it easier or harder, depending on your performance on each trial. It's set up so you should only be able to stop about HALF of the time, so you should NOT feel like you are making a serious error if you can't stop on some trials. So, you should just try to go as quickly as you can without making mistakes, and really try to stop if you can when you hear a beep, but know that you won't be able to stop on about half the trials.

Because the program adjusts to how well you are doing, it won't help to slow down and try to wait for the beep, though it's natural to slow down a little. So please keep trying to respond as quickly and accurately as you can. At the end of each block, there will be feedback about your response time and the % of time you stopped, so I'll be able to tell if you are following the instructions.

Now there will be 3 blocks of real trials. You will be doing the same exact thing as in the practice trials, and you can take little breaks in between the blocks if you need. Do you have any questions before you begin?

<u>Things to watch for in RTs and accuracy:</u>

- If their probability of stopping is around 50% (40% to 60% is ok), they are doing a great job.
    - If any of their blocks falls outside 40-60% range, make a note of it on the Notes Page
- Sometimes the RT will be fast but they won't be able to stop that much. In this case, tell them they are doing a good job with the RT, so keep at that, but do their best to stop when they can. There aren't that many trials in the practice, so you can tell them that it may be hard to get used to it.
- If they never stop, tell them you are going to give them another practice so they can try to stop (hit esc twice, then select test again; make a note that you had to do this).

<u>Things to emphasize:</u>

- We want their performance to be fairly CONSISTENT across blocks. So if they start doing something weird like waiting for the beeps (slow WAY down) in the 2nd block or stop stopping, etc., that's a problem – <u>remind them that they should only be able to stop 50% of the time, and it's okay to press a button after hearing a beep</u>.
- Go over their % stopping after each block. If they are not following instructions, remind them they should only be able to stop 50% of the time.

- The end info also gives their accuracy (Incorrect Go RT xx%), which should be high. If they are just hitting one button or doing something strange you'll be able to tell that the accuracy is very low.

************** OPTIONAL 5 MINUTE BREAK **************

Priming

- Have participant sit in front of PC.
- Have participant put on headphones and adjust sound to a comfortable level. They will not able to adjust the sound once the task begins.
- During the task, make sure the participant is hitting the correct buttons (m or z). If they are not clicking on the correct keys or not responding, tap them on the shoulder and remind them to answer as quickly and as accurately as possible.

Open one of the lists (refer to the participant tracking sheet):

1. Go to the folder **Desktop\Alia\diss_exp**
2. If PsychoPy is not already open, open the software by clicking on **PsychoPy2** icon (a circle with black and white lines)
3. Select *File* -> *Open*
1. A window will pop up asking you which list to open.
    a. Navigate to **Desktop\Alia\diss_exp\priming**
    b. Choose either **pp_a.psyexp** or **pp_b.psyexp** depending on what is listed on the participant tracking sheet.
2. In the newly opened window, click the *run* button, which is the green button with a running person in the middle
3. Enter the subject number and push enter

Provide instructions:
In this task, you will hear pairs of words on via the headphones. Your task will be to decide if the second word in each pair is a real word in English (like blanket) or not (like blicket). Press M if you think the second word is a real word and Z if you think it is not. After every trial, you will receive feedback on your accuracy and how fast your response was.

Pursuit rotor

- Have participant sit in front of PC. They will not need the headphones for this task
- The cursor just has to be on the red dot. The mouse can be clicked or not.

Open the pursuit rotor task on the PEBL software

1. Go to the folder **Desktop\Alia\diss_exp**
2. Click on the **pursuitrotor** shortcut file to start the task.
3. Enter the subject number and push enter

Provide instructions:

In this task, you will see a circle in the middle of the screen and a red dot that will travel on the outside of the circle. Your task is to keep the mouse cursor on the red dot as it moves around the circle. You'll do this 4 times.

************** OPTIONAL 5 MINUTE BREAK **************

Stroop

- Have participant sit in front of PC. They will not need headphones for this task.
- The task changes periodically, so you'll need to make sure they understand the instructions.
- Occasionally check that the subject is not cheating by looking away from the stimuli.

Open the task in PsychoPy:

4. Go to the folder **Desktop\Alia\diss_exp**
5. If PsychoPy is not already open, open the software by clicking on **PsychoPy2** icon (a circle with black and white lines)
6. Select *File* -> *Open*
4. A window will pop up asking you which list to open.
    a. Navigate to **Desktop\Alia\diss_exp\stroop**
    b. Choose the **stroop_buttonpress.psyexp** file
5. In the newly opened window, click the *run* button, which is the green button with a running person in the middle
6. Enter the subject number and push enter

Provide instructions:
In the next task, you pressing the left, down, and right arrow buttons to indicate the color you see on the screen. Please read through the instructions and let me know if you have any questions.

Let's begin with some practice trials.

Antisaccade

- Have participant sit in front of Mac.
- MAKE SURE THE SUBJECT IS 18 INCHES FROM THE SCREEN. Use the string behind the computer to measure the distance, and tell the subject to make sure he/she is comfortable and not to lean forward or backward. Move screen rather than person moving. Make sure the computer screen is centered.
- Make sure the keyboard is in front of you so you can type in the numbers the participant says.
- Re-measure the distance if they begin to drift.
- Subtly watch their eyes from time to time (1+ times/block) to be sure they are saccading or antisaccading when appropriate. This task is demanding and people like to try to move their eyes early, anticipating which side the cue will appear on or they may stop saccading and try to determine the number only using peripheral vision. You must keep them honest by monitoring that they are always staying

focused on the fixation *until* the cue flashes at which time they must saccade in one direction or another. If they begin to cheat, then write down the next number they tell you (so you can type it in after you speak to them) and then say: “Remember to keep your eyes on the fixation until the cue flashes.” Or whatever reminder is appropriate to keep them honest.

- As the experimenter, it’s better not to watch the screen throughout, because it makes your task easier on those incorrect trials where the number they say does not match the number on the screen.

Open the task in PsychScope:

1. Go to the folder: **Desktop\experiments\alia_diss**
2. Click on the **Antisaccade.command** folder to start the recording and open the task in PsychScope
3. Click Enter to start the recording, and enter the ID number when prompted
4. Click on the “Run” menu and select “Run”
5. Enter the participant number
6. Save the data in: **Desktop\experiments\alia_diss\Antisaccade\data**
7. At the end of the task, press enter to stop the recording

Provide instructions:
In this task, you will be looking at a fixation on the screen and then you will move your eyes away as soon as the cue flashes. You will then say the number that you see out loud and I will type it in on the keyboard. We will be recording what you say with the internal microphone in case we need to double check what we’ve typed in, so make sure to speak loud enough so that recording can be heard.
Make sure you don’t lean forward or backward throughout the task. We want to make sure that you stay at the same distance from the monitor at all times. Please read through the instructions on the screen. (wait while they read the instructions)

Familiarity

- **Only for L2 speakers**
- Have participant sit in front of PC. They will not need headphones for this task.
- During the task, make sure the participant is providing a translation and rating for each word (not just typing the same word and giving the same rating for each word).

Open the task in PsychoPy:

7. Go to the folder **Desktop\Alia\diss_exp**
8. If PsychoPy is not already open, open the software by clicking on **PsychoPy2** icon (a circle with black and white lines)
9. Select *File* -> *Open*
7. A window will pop up asking you which list to open.
    a. Navigate to **Desktop\Alia\diss_exp\familiarity**
    b. Choose the **familiarity.psyexp** file
8. In the newly opened window, click the *run* button, which is the green button with a running person in the middle

9. Enter the subject number and push enter

Provide instructions:
In this task, you will see a series of English words. For each word, you will do 2 things: 1) type a Spanish translation, and 2) rate how familiar you are with the word. For the Spanish translation, don't worry too much about spelling and do NOT include accents or articles (such as el). Type in the first translation that comes to mind, and if you don't know one, type your best guess. After providing the translation, you will rate the English word on a scale of 1 (I have never seen this word before) to 5 (I know this word very well).

There will be some practice trials before you begin.

Final Payment form

- **Only for paid participants**
- Give participant individual receipt and payment log

Appendix D: Visual-world stimuli

The stimuli in Table 25 list the pairs of words displayed during the visual-world task, experimental (i.e., numbered pair) and filler trials. Words in a pair were tagged as higher or lower in frequency, according to the SUBTLEX corpus (Brysbaert & New, 2009). Table 26 provides the experimental pair characteristics, such as cognate status and syllable length similarity. List composition balanced target and competitor status, such that each word in a pair functioned as one or the other across participants.

Table 25: Visual-world stimuli lexical characteristics

Word	Pair	Relative Frequency	Frequency pm	Frequency log	Neighbors	Neighborhood Frequency log
crown	1	control	13.69	2.8445	11	14.9322
belt	1	HF	24.35	3.0945	18	47.2255
bench	1	LF	9.67	2.6937	9	22.5055
peanut	2	control	12.35	2.8000	1	7.5490
bridge	2	HF	45.71	3.3677	9	40.1111
brick	2	LF	10.18	2.7160	20	42.6882
fence	3	control	16.06	2.9138	12	14.5948
bread	3	HF	28.33	3.1602	19	28.0279
branch	3	LF	10.08	2.7118	6	8.0686
ghost	4	control	36.59	3.2711	14	38.7815
brain	4	HF	77.02	3.5943	21	26.4164
bride	4	LF	24.22	3.0920	22	23.6542
sailor	5	control	12.39	2.8014	17	5.8662
butter	5	HF	20.43	3.0183	26	205.2805
bubble	5	LF	8.00	2.6117	12	8.6879
table	6	control	105.63	3.7314	12	21.1160
doctor	6	HF	263.94	4.1291	2	18.7353
dollar	6	LF	27.65	3.1495	10	15.0157
wallet	7	control	22.80	3.0660	3	6.5098
desert	7	HF	27.98	3.1547	0	0.0000
dentist	7	LF	11.20	2.7574	1	1.0392
books	8	control	67.76	3.5387	16	42.6164
dress	8	HF	87.20	3.6482	9	19.0763
drill	8	LF	13.75	2.8463	14	2.0798
ticket	9	control	45.57	3.3664	4	11.7794

Word	Pair	Relative Frequency	Frequency pm	Frequency log	Neighbors	Neighborhood Frequency log
film	9	HF	65.25	3.5223	10	13.3451
finger	9	LF	36.67	3.2721	4	8.9951
sink	10	control	16.92	2.9365	27	114.1990
frame	10	HF	14.10	2.8573	12	173.4444
frog	10	LF	11.82	2.7810	5	3.3137
suitcase	11	control	13.39	2.8351	0	0.0000
glass	11	HF	60.71	3.4909	9	42.7799
glove	11	LF	10.10	2.7126	6	189.1797
basketball	12	control	21.39	3.0382	0	0.0000
hand	12	HF	279.65	4.1542	24	651.3194
hammer	12	LF	12.47	2.8041	15	2.2536
jacket	13	control	33.41	3.2317	5	6.6981
camera	13	HF	57.00	3.4636	1	17.9412
cabin	13	LF	19.65	3.0013	1	0.9412
shoulder	14	control	26.20	3.1261	7	8.8515
candy	14	HF	35.78	3.2615	10	6.8196
candle	14	LF	8.02	2.6128	9	20.8170
lips	15	control	31.18	3.2017	19	4.3468
clock	15	HF	58.63	3.4758	21	8.1634
clown	15	LF	15.82	2.9074	9	19.4924
flag	16	control	17.49	2.9509	11	5.4546
closet	16	HF	27.08	3.1405	1	2.1373
cloud	16	LF	11.75	2.7782	12	14.0065
glasses	17	control	33.12	3.2279	4	4.5098
queen	17	HF	54.69	3.4456	5	27.7530
quarter	17	LF	26.02	3.1232	6	3.6209
lightning	18	control	14.14	2.8585	1	0.7647
palace	18	HF	19.20	2.9912	3	32.0523
pancakes	18	LF	9.65	2.6928	1	3.9608
toilet	19	control	28.90	3.1688	1	3.3333
pants	19	HF	58.75	3.4767	10	5.7490
package	19	LF	22.78	3.0656	3	3.5556
basket	20	control	13.18	2.8280	3	1.8432
pocket	20	HF	35.71	3.2605	7	4.7647
popcorn	20	LF	9.12	2.6684	0	0.0000
cookie	21	control	16.71	2.9309	14	5.9860
penny	21	HF	24.29	3.0934	16	98.8076
pencil	21	LF	9.86	2.7024	4	1.1569
letter	22	control	82.61	3.6247	20	180.4902
picture	22	HF	138.45	3.8489	4	18.6618
pillow	22	LF	11.39	2.7649	7	11.3081
cowboy	23	control	18.98	2.9863	1	4.5294

Word	Pair	Relative Frequency	Frequency pm	Frequency log	Neighbors	Neighborhood Frequency log
square	23	HF	31.76	3.2098	5	25.6196
skull	23	LF	14.71	2.8756	10	36.5686
grass	24	control	16.78	2.9330	16	18.4449
stairs	24	HF	23.76	3.0839	10	10.0471
sticks	24	LF	13.61	2.8420	15	22.6314
forest	25	control	18.88	2.9841	2	2.0000
stone	25	HF	40.63	3.3166	18	6.9662
steak	25	LF	16.24	2.9186	21	139.2502
files	26	control	26.63	3.1332	17	16.6021
street	26	HF	148.18	3.8784	7	31.7395
string	26	LF	12.67	2.8109	12	18.0131
lion	27	control	15.35	2.8943	7	32.1961
taxi	27	HF	25.84	3.1202	5	3.8196
tattoo	27	LF	11.96	2.7860	5	1.8510
snake	28	control	22.35	3.0573	8	17.4461
trash	28	HF	22.47	3.0596	9	13.5861
trail	28	LF	19.20	2.9912	18	15.4684
target	29	control	37.96	3.2871	1	6.6667
truck	29	HF	72.86	3.5702	8	20.3162
trunk	29	LF	19.80	3.0048	4	38.3676
tent	30	control	17.49	2.9509	22	45.2398
chest	30	HF	40.98	3.3204	20	46.0706
cherry	30	LF	13.59	2.8414	11	129.3672
mirror	31	control	24.18	3.0913	2	13.7255
chicken	31	HF	61.73	3.4982	2	5.6274
chips	31	LF	16.24	2.9186	17	7.0346
rabbit	32	control	20.94	3.0290	3	7.1961
window	32	HF	86.00	3.6422	5	20.9921
whistle	32	LF	15.45	2.8971	10	214.7765
box	33	control	89.75	3.6607	29	9.0690
woman	33	HF	434.63	4.3457	3	2.7059
wolf	33	LF	20.27	3.0149	2	1.8137
barrel	filler	comp	10.63	2.7348	7	9.8291
chocolate	filler	target	29.39	3.1761	1	2.9216
lipstick	filler	target	8.80	2.6532	1	0.4902
newspaper	filler	target	23.69	3.0824	1	11.4706
music	filler	comp	151.65	3.8885	2	7.5981
robot	filler	comp	12.18	2.7938	2	2.7843
motorcycl	filler	comp	8.92	2.6590	1	1.2941
diamond	filler	target	20.65	3.0228	1	17.0392
treasure	filler	comp	19.06	2.9881	3	1.3203
skirt	filler	target	9.96	2.7067	12	5.8448

Word	Pair	Relative Frequency	Frequency pm	Frequency log	Neighbors	Neighborhood Frequency log
garden	filler	comp	26.55	3.1319	4	18.3579
sweater	filler	target	13.80	2.8482	7	4.7731
shadow	filler	comp	21.18	3.0338	6	3.7615
balloon	filler	target	8.67	2.6464	2	4.9118
scale	filler	comp	9.51	2.6866	12	34.0735
twins	filler	target	15.10	2.8871	5	6.3176
wagon	filler	comp	17.76	2.9576	1	3.2353
angel	filler	target	78.27	3.6013	2	8.6471
socks	filler	comp	18.27	2.9699	29	20.5030
ladder	filler	target	9.25	2.6749	14	5.4104
policeman	filler	comp	11.73	2.7774	1	3.8627
nest	filler	target	11.10	2.7536	21	64.4790
necklace	filler	comp	9.75	2.6972	2	3.0196
refrigerato	filler	target	8.37	2.6314	1	0.4902
helmet	filler	comp	9.47	2.6848	1	2.0196
priest	filler	target	26.20	3.1261	5	2.4745
barbecue	filler	comp	8.94	2.6599	3	0.6405
statue	filler	target	10.59	2.7332	1	1.9412
envelope	filler	comp	10.06	2.7110	0	0.0000
waiter	filler	target	13.20	2.8287	18	78.3475
drum	filler	comp	8.47	2.6365	9	253.8911
needle	filler	target	11.92	2.7846	7	210.8431
doorbell	filler	comp	8.33	2.6294	0	0.0000
banana	filler	target	10.73	2.7388	1	5.4510

Notes. *pm* indicates instances per million in the corpus.

Table 26: Visual-world stimuli pair characteristics

Pair	Contains Confusable Phoneme	Same Syllable Length	Overlap Type	Contains Cognates	Low Name Agreement
1	no	yes	CV	no	no
2	no	yes	CCV	no	no
3	yes	yes	CC	no	no
4	yes	yes	CC	no	no
5	no	yes	CV	no	no
6	no	yes	CV	both	yes
7	no	yes	CV	both	no
8	yes	yes	CC	no	no
9	no	no	CV	no	yes
10	no	yes	CC	no	no
11	yes	yes	CC	no	no
12	no	no	CV	no	no
13	no	yes	CV	both	no
14	no	yes	CV	no	no
15	no	yes	CC	no	no
16	no	no	CC	no	no
17	no	no	CC	no	no
18	no	yes	CV	both	yes
19	no	no	CV	both	no
20	no	yes	CV	no	no
21	no	yes	CVC	no	no
22	no	yes	CVC	no	no
23	yes	yes	CC	no	no
24	no	yes	CC	no	no
25	yes	yes	CC	no	yes
26	no	yes	CCV	no	yes
27	no	yes	CV	both	no
28	no	yes	CC	no	yes
29	no	yes	CCV	no	yes
30	no	no	CV	no	no
31	no	no	CV	no	yes
32	no	yes	CV	no	no
33	no	no	CV	no	no

Appendix E: Phonological priming stimuli

Table 27 displays the experimental (aka exp) and filler trial stimuli for the phonological priming task. Relatedness varied across list.

Table 27: Phonological priming lexical and pair characteristics

Word	Pair	Word Type	Pair Type	Frequency pm	Frequency log	Neighbors	Overlap Type
outlet	1	prime	exp	2.00	2.0128	1	VCC
outline	1	target	exp	2.04	2.0212	2	VCC
advantages	2	prime	exp	2.82	2.1614	0	VCC
advisor	2	target	exp	4.41	2.3541	3	VCC
freely	3	prime	exp	4.22	2.3345	2	CCV
freezer	3	target	exp	5.16	2.4216	2	CCV
stale	4	prime	exp	2.92	2.1761	24	CCV
stain	4	target	exp	6.20	2.5011	20	CCV
brag	5	prime	exp	3.51	2.2553	11	CCV
brat	5	target	exp	6.22	2.5024	18	CCV
reply	6	prime	exp	4.80	2.3909	3	CVCV
replacement	6	target	exp	6.22	2.5024	1	CVCV
scramble	7	prime	exp	2.41	2.0934	3	CCC
scrub	7	target	exp	6.24	2.5038	2	CCC
groceries	8	prime	exp	5.90	2.4800	1	CCV
growth	8	target	exp	6.45	2.5185	7	CCV
blankets	9	prime	exp	5.10	2.4166	1	CCV
blackmail	9	target	exp	7.73	2.5966	2	CCV
complain	10	prime	exp	12.55	2.8069	3	CVCC
command	10	target	exp	43.63	3.3475	4	CVCC
deserved	11	prime	exp	10.06	2.7110	4	CVC
disturb	11	target	exp	11.10	2.7536	2	CVC
kidney	12	prime	exp	9.69	2.6946	4	CVCC
kidnapped	12	target	exp	16.39	2.9227	2	CVCC
snack	13	prime	exp	9.14	2.6693	13	CCV
snap	13	target	exp	17.39	2.9484	13	CCV
reverse	14	prime	exp	10.98	2.7490	4	CVC
revenge	14	target	exp	19.04	2.9877	0	CVC
breath	15	prime	exp	44.92	3.3602	8	CCV
breakfast	15	target	exp	66.29	3.5292	1	CCV
sends	16	prime	exp	11.47	2.7679	9	CVC
sentence	16	target	exp	20.53	3.0204	1	CVC
squeeze	17	prime	exp	15.08	2.8865	7	CCC

Word	Pair	Word Type	Pair Type	Frequency pm	Frequency log	Neighbors	Overlap Type
squad	17	target	exp	21.49	3.0402	8	CCC
prior	18	prime	exp	8.27	2.6263	13	CCV
pride	18	target	exp	27.67	3.1498	19	CCV
delay	19	prime	exp	11.02	2.7505	5	CVC
deliver	19	target	exp	28.35	3.1605	2	CVC
crash	20	prime	exp	28.65	3.1649	15	CCV
crack	20	target	exp	32.84	3.2243	24	CCV
remain	21	prime	exp	33.22	3.2292	5	CVC
remind	21	target	exp	36.92	3.2751	6	CVC
scream	22	prime	exp	26.41	3.1297	9	CCC
screw	22	target	exp	37.49	3.2817	6	CCC
selfish	23	prime	exp	15.90	2.9096	1	CVC
selling	23	target	exp	37.63	3.2833	13	CVC
brave	24	prime	exp	31.71	3.2090	12	CCV
breaking	24	target	exp	42.25	3.3336	3	CCV
health	25	prime	exp	40.27	3.3128	9	CVC
held	25	target	exp	42.45	3.3357	18	CVC
dummy	26	prime	exp	9.80	2.6998	13	CVC
dump	26	target	exp	28.82	3.1676	17	CVC
below	27	prime	exp	28.04	3.1556	16	CVC
build	27	target	exp	48.08	3.3897	21	CVC
player	28	prime	exp	37.76	3.2849	14	CCV
places	28	target	exp	53.06	3.4325	2	CCV
dizzy	29	prime	exp	8.43	2.6345	10	CVC
disease	29	target	exp	26.18	3.1258	2	CVC
travel	30	prime	exp	33.37	3.2312	5	CCV
track	30	target	exp	55.75	3.4539	14	CCV
heavy	31	prime	exp	47.29	3.3826	5	CVC
heaven	31	target	exp	56.61	3.4606	5	CVC
delighted	32	prime	exp	12.94	2.8202	1	CVC
delivery	32	target	exp	19.94	3.0077	1	CVC
happen	33	prime	exp	254.27	4.1129	2	CVC
happy	33	target	exp	333.20	4.2303	10	CVC
fired	34	prime	exp	61.94	3.4997	11	CVC
fighting	34	target	exp	70.80	3.5577	10	CVC
quit	35	prime	exp	90.10	3.6624	19	CCV
quick	35	target	exp	108.67	3.7437	14	CCV
fault	36	prime	exp	104.12	3.7252	14	CVC
follow	36	target	exp	123.20	3.7982	13	CVC
spit	37	prime	exp	14.63	2.8733	14	CCV
spin	37	target	exp	19.35	2.9948	16	CCV
trust	38	prime	exp	178.18	3.9585	8	CCV

Word	Pair	Word Type	Pair Type	Frequency pm	Frequency log	Neighbors	Overlap Type
trouble	38	target	exp	223.55	4.0570	5	CCV
speed	39	prime	exp	41.25	3.3233	16	CCV
speaking	39	target	exp	69.90	3.5522	7	CCV
station	40	prime	exp	79.08	3.6057	3	CCVC
state	40	target	exp	107.84	3.7404	19	CCVC
ceiling	f1	prime	filler	8.35	2.6304	18	-
shortly	f1	target	filler	10.06	2.7110	1	-
succeed	f2	prime	filler	9.45	2.6839	1	-
dessert	f2	target	filler	14.02	2.8549	2	-
haircut	f3	prime	filler	8.41	2.6335	1	-
joking	f3	target	filler	18.61	2.9777	5	-
slide	f4	prime	filler	17.82	2.9590	17	-
bones	f4	target	filler	30.61	3.1937	22	-
failure	f5	prime	filler	20.02	3.0095	1	-
coffin	f5	target	filler	9.04	2.6646	4	-
speech	f6	prime	filler	38.04	3.2880	4	-
closer	f6	target	filler	45.67	3.3674	3	-
study	f8	prime	filler	49.04	3.3983	6	-
blame	f8	target	filler	58.78	3.4770	9	-
danger	f7	prime	filler	43.67	3.3479	4	-
silly	f7	target	filler	57.10	3.4643	25	-
careful	f9	prime	filler	108.82	3.7444	0	-
evening	f9	target	filler	120.69	3.7893	0	-
ankle	f10	prime	filler	8.02	2.6128	3	-
spider	f10	target	filler	10.10	2.7126	4	-
smack	f11	prime	filler	9.51	2.6866	13	-
bacon	f11	target	filler	11.86	2.7825	5	-
safely	f12	prime	filler	11.10	2.7536	1	-
punishment	f12	target	filler	13.43	2.8363	1	-
stink	f13	prime	filler	13.20	2.8287	12	-
brand	f13	target	filler	13.96	2.8531	14	-
cotton	f14	prime	filler	14.18	2.8597	6	-
dancer	f14	target	filler	16.29	2.9201	6	-
device	f15	prime	filler	18.16	2.9671	4	-
amount	f15	target	filler	24.75	3.1014	3	-
approach	f16	prime	filler	20.98	3.0298	1	-
returned	f16	target	filler	24.76	3.1017	3	-
library	f17	prime	filler	22.94	3.0686	1	-
stole	f17	target	filler	53.16	3.4333	18	-
friendly	f18	prime	filler	26.04	3.1235	0	-
weather	f18	target	filler	34.24	3.2423	12	-
stuck	f19	prime	filler	66.65	3.5315	20	-

Word	Pair	Word Type	Pair Type	Frequency pm	Frequency log	Neighbors	Overlap Type
field	f19	target	filler	70.20	3.5540	22	-
test	f20	prime	filler	84.08	3.6324	22	-
strong	f20	target	filler	86.86	3.6465	5	-
arrive	f21	prime	nonword	18.69	2.9795	8	-
voarsely	f21	target	nonword				-
fifth	f22	prime	nonword	19.20	2.9912	1	-
nollow	f22	target	nonword				-
argue	f23	prime	nonword	19.75	3.0035	2	-
thance	f23	target	nonword				-
anger	f24	prime	nonword	19.43	2.9965	6	-
erpand	f24	target	nonword				-
leads	f25	prime	nonword	19.78	3.0043	20	-
banser	f25	target	nonword				-
hidden	f26	prime	nonword	21.27	3.0358	2	-
pateful	f26	target	nonword				-
deny	f27	prime	nonword	21.39	3.0382	4	-
frain	f27	target	nonword				-
refuse	f28	prime	nonword	20.98	3.0298	6	-
bealm	f28	target	nonword				-
plain	f29	prime	nonword	21.82	3.0469	22	-
apea	f29	target	nonword				-
range	f30	prime	nonword	22.76	3.0652	15	-
melp	f30	target	nonword				-
studio	f31	prime	nonword	23.33	3.0759	1	-
mons	f31	target	nonword				-
money	f32	prime	nonword	640.76	4.5143	20	-
enector	f32	target	nonword				-
mostly	f33	prime	nonword	26.08	3.1242	1	-
baddock	f33	target	nonword				-
slip	f34	prime	nonword	25.88	3.1209	20	-
beon	f34	target	nonword				-
grave	f35	prime	nonword	26.27	3.1274	19	-
bram	f35	target	nonword				-
sports	f36	prime	nonword	27.59	3.1486	10	-
trake	f36	target	nonword				-
switch	f37	prime	nonword	28.12	3.1569	14	-
mediul	f37	target	nonword				-
winner	f38	prime	nonword	31.22	3.2022	17	-
felm	f38	target	nonword				-
season	f39	prime	nonword	31.47	3.2057	7	-
balbing	f39	target	nonword				-
handsome	f40	prime	nonword	33.02	3.2266	0	-

Word	Pair	Word Type	Pair Type	Frequency pm	Frequency log	Neighbors	Overlap Type
fitterbug	f40	target	nonword				
yellow	f41	prime	nonword	33.80	3.2368	10	-
glaight	f41	target	nonword				-
freak	f42	prime	nonword	36.75	3.2730	14	-
drock	f42	target	nonword				-
bound	f43	prime	nonword	18.43	2.9736	18	-
fristle	f43	target	nonword				-
pretend	f44	prime	nonword	40.31	3.3132	2	-
dovet	f44	target	nonword				-
afford	f45	prime	nonword	44.43	3.3555	6	-
prain	f45	target	nonword				-
blind	f46	prime	nonword	45.82	3.3688	9	-
bunfle	f46	target	nonword				-
dressed	f47	prime	nonword	46.94	3.3793	9	-
lassan	f47	target	nonword				-
depth	f48	prime	nonword	8.25	2.6253	2	-
dift	f48	target	nonword				-
waste	f49	prime	nonword	53.25	3.4341	19	-
curp	f49	target	nonword				-
flight	f50	prime	nonword	59.69	3.4836	20	-
braction	f50	target	nonword				-
forward	f51	prime	nonword	72.33	3.5670	1	-
tranch	f51	target	nonword				-
smell	f52	prime	nonword	83.14	3.6275	12	-
dacy	f52	target	nonword				-
fast	f53	prime	nonword	137.45	3.8458	20	-
bloot	f53	target	nonword				-
fact	f54	prime	nonword	172.57	3.9446	21	-
curtaip	f54	target	nonword				-
straight	f55	prime	nonword	122.43	3.7955	11	-
sauge	f55	target	nonword				-
display	f56	prime	nonword	8.53	2.6395	3	-
leity	f56	target	nonword				-
journal	f57	prime	nonword	8.88	2.6571	5	-
gomely	f57	target	nonword				-
fairly	f58	prime	nonword	9.18	2.6712	10	-
giesta	f58	target	nonword				-
grip	f59	prime	nonword	9.69	2.6946	16	-
carm	f59	target	nonword				-
bounce	f60	prime	nonword	9.84	2.7016	7	-
bont	f60	target	nonword				-
plug	f61	prime	nonword	10.41	2.7259	11	-

Word	Pair	Word Type	Pair Type	Frequency pm	Frequency log	Neighbors	Overlap Type
flosh	f61	target	nonword				-
laughed	f62	prime	nonword	10.69	2.7372	4	-
interlade	f62	target	nonword				-
pumpkin	f63	prime	nonword	10.84	2.7435	2	-
noney	f63	target	nonword				-
carpet	f64	prime	nonword	11.65	2.7745	1	-
sierce	f64	target	nonword				-
blonde	f65	prime	nonword	13.92	2.8519	7	-
nilt	f65	target	nonword				-
rocket	f66	prime	nonword	11.84	2.7818	7	-
plurry	f66	target	nonword				-
link	f67	prime	nonword	11.94	2.7853	20	-
cessian	f67	target	nonword				-
sharing	f68	prime	nonword	12.22	2.7952	3	-
fadler	f68	target	nonword				-
fries	f69	prime	nonword	11.69	2.7760	16	-
phafe	f69	target	nonword				-
railroad	f70	prime	nonword	12.43	2.8028	1	-
drail	f70	target	nonword				-
steam	f71	prime	nonword	13.45	2.8370	17	-
carmer	f71	target	nonword				-
desk	f72	prime	nonword	43.90	3.3502	6	-
timiny	f72	target	nonword				-
contest	f73	prime	nonword	18.78	2.9818	4	-
scade	f73	target	nonword				-
stretch	f74	prime	nonword	14.67	2.8745	5	-
solly	f74	target	nonword				-
salad	f75	prime	nonword	17.02	2.9390	2	-
connod	f75	target	nonword				-
jumping	f76	prime	nonword	14.27	2.8627	5	-
marfy	f76	target	nonword				-
fairy	f77	prime	nonword	16.69	2.9304	13	-
crisly	f77	target	nonword				-
neighbor	f78	prime	nonword	16.94	2.9370	5	-
bloatev	f78	target	nonword				-
highway	f79	prime	nonword	17.86	2.9600	1	-
abert	f79	target	nonword				-
eleven	f80	prime	nonword	12.98	2.8215	1	-
bation	f80	target	nonword				

Note. exp indicates an experimental world or trial

Appendix F: Full model output

Visual world RQ1

Visual world full multilevel model output for all speakers comparing conditions (unrelated, related).

Table 28: Sum coding applied to condition model coefficients for all visual-world outcome multilevel model estimates

Condition	Contrast codes
Unrelated	1
Related	-1

Table 29: Multilevel model log transformed RT estimates for all speakers

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	7.118	0.017	104.900	407.385	<0.001
condition1	-0.013	0.003	2513.000	-3.799	<0.001
Random effects	Variance	*SD*	Correlation		
Subject (Intercept)	0.017	0.129			
condition1	<0.001	<0.001	1.000		
Item (Intercept)	0.003	0.055			
Residual	0.029	0.170			

Table 30: Multilevel model *z*-score AUC estimates for all speakers

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	0.044	0.030	31.930	1.468	0.152
condition1	-0.121	0.020	10.320	-6.006	<0.001
Random effects	Variance	*SD*	Correlation		
Subject (Intercept)	<0.001	<0.001			
condition1	0.001	0.030			
Item (Intercept)	0.017	0.130			
Residual	1.044	1.022			

Table 31: Multilevel model z-score MD estimates for all speakers

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	0.048	0.029	32.120	1.680	0.103
condition1	-0.111	0.021	86.680	-5.263	<0.001
Random effects	Variance	*SD*	Correlation		
Subject (Intercept)	<0.001	0.015			
condition1	0.004	0.064	-1.000		
Item (Intercept)	0.014	0.120			
Residual	1.039	1.020			

Table 32: Multilevel model *x*-axis flips (Poisson distribution) estimates for all speakers

Fixed effects	*b*	*SE*	*z*-value	*p*-value
(Intercept)	1.908	0.027	71.610	<0.001
condition1	-0.007	0.009	-0.770	0.440
Random effects	Variance	*SD*	Correlation	
Subject (Intercept)	0.053	0.230		
condition1	0.002	0.044	0.430	
Item (Intercept)	0.000	0.022		

Visual world RQ3 all speakers

Visual world full multilevel model output for all speakers in related conditions. IC composite was centered.

Table 33: Sum coding applied to speaker type model coefficients for all outcome multilevel model estimates

Condition	Contrast codes
Monolingual	1
Bilingual	-1

Table 34: Multilevel model log transformed RT estimates for all speakers in related conditions

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	7.135	0.018	89.400	386.651	0.000
spkr1	-0.059	0.015	77.900	-4.004	0.000
IC composite	0.000	0.015	79.600	0.009	0.993
Frequency difference	-0.033	0.008	1142.000	-4.206	0.000
spkr1:IC composite	-0.005	0.015	81.800	-0.343	0.733
spkr1:Frequency difference	0.008	0.008	1145.500	1.053	0.292
IC composite:Frequency difference	-0.011	0.008	1143.800	-1.431	0.153
spkr1:IC composite:Frequency difference	0.007	0.008	1091.600	0.933	0.351
Random effects	**Variance**	***SD***	**Correlation**		
Subject (Intercept)	0.014	0.117			
Frequency difference	<0.001	0.002	-1.000		
Item (Intercept)	0.004	0.064			
spkr1	<0.001	0.010	-0.350		
IC composite	<0.001	0.014	0.330	-0.950	
spkr1:IC composite	<0.001	0.020	-0.090	0.890	-0.970
Residual	0.026	0.162			

Table 35: Multilevel model z-scored AUC estimates for all speakers in related conditions

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	0.16	0.06	32.2	2.6	0.014
spkr1	0.038	0.041	37.900	0.931	0.358
IC composite	-0.015	0.034	411.200	-0.436	0.663
Frequency difference	-0.056	0.054	1254.700	-1.035	0.301
spkr1:IC composite	0.016	0.034	314.300	0.477	0.634
spkr1:Frequency difference	0.024	0.054	1256.100	0.451	0.652
IC composite:Frequency difference	0.005	0.053	1260.700	0.102	0.919
spkr1:IC composite: Frequency difference	0.037	0.053	1258.4	0.693	0.489
Random effects	**Variance**	***SD***	**Correlation**		
Subject (Intercept)	<0.001	<0.001			
Frequency difference	<0.001	<0.001			
Item (Intercept)	0.086	0.293			
spkr1	0.017	0.129	0.900		
IC composite	0.001	0.027	0.360	0.740	
spkr1:IC composite	0.001	0.031	-0.330	0.120	0.760
Residual	1.299	1.140			

Table 36: Multilevel model z-scored MD estimates for all speakers in related conditions

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	0.158	0.060	32.200	2.651	0.012
spkr1	0.015	0.038	41.800	0.393	0.697
IC composite	-0.002	0.034	326.700	-0.049	0.961
Frequency difference	-0.055	0.052	1268.500	-1.042	0.298
spkr1:IC composite	0.012	0.034	192.400	0.348	0.728
spkr1:Frequency difference	0.011	0.053	1264.600	0.217	0.828
IC composite:Frequency difference	0.024	0.052	1264.900	0.470	0.638
spkr1:IC composite: Frequency difference	0.042	0.051	1271.700	0.820	0.412
Random effects	Variance	*SD*	Correlation		
Subject (Intercept)	<0.001	<0.001			
Frequency difference	<0.001	<0.001			
Item (Intercept)	0.081	0.284			
spkr1	0.012	0.111	1.000		
IC composite	0.001	0.033	0.120	0.120	
spkr1:IC composite	0.002	0.039	0.070	0.070	
Residual	1.227	1.108			

Table 37: Multilevel model *x*-axis flips (Poisson distribution) estimates for all speakers in related conditions

Fixed effects	*b*	*SE*	*z*-value	*p*-value
(Intercept)	1.923	0.028	68.880	<0.001
spkr1	-0.004	0.028	-0.140	0.891
IC composite	-0.008	0.028	-0.270	0.786
Frequency difference	-0.024	0.019	-1.240	0.215
spkr1:IC composite	-0.025	0.029	-0.880	0.379
spkr1:Frequency difference	0.000	0.019	0.010	0.994
IC composite:Frequency difference	-0.014	0.018	-0.780	0.436
spkr1:IC composite: Frequency difference	0.022	0.018	1.180	0.240
Random effects	**Variance**	***SD***	**Correlation**	
Subject (Intercept)	0.045	0.212		
Frequency difference	0.002	0.039	0.290	
Item (Intercept)	0.001	0.029		
spkr1	<0.001	0.018	-1.000	
IC composite	<0.001	0.009	1.000	-1.000
spkr1:IC composite	0.002	0.041	-1.000	1.000

Visual world RQ3 monolinguals

Visual world full multilevel model output for bilinguals in related conditions. IC composite was centered.

Table 38: Multilevel model log transformed RT estimates for monolinguals in related conditions

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	7.075	0.020	58.600	354.816	<0.001
IC composite	-0.005	0.018	40.400	-0.255	0.800
Frequency difference	-0.024	0.010	515.200	-2.483	0.013
IC composite:Frequency difference	-0.006	0.011	518.000	-0.526	0.599
Random effects	**Variance**	***SD***	**Correlation**		
Subject (Intercept)	0.008	0.092			
Frequency difference	<0.001	0.005	-1.000		
Item (Intercept)	0.004	0.063			
IC composite	<0.001	0.008	0.350		
Residual	0.022	0.147			

Table 39: Multilevel model *z*-scored AUC estimates for monolinguals in related conditions

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	0.197	0.078	33.300	2.514	0.017
IC composite	0.001	0.054	391.800	0.027	0.978
Frequency difference	-0.031	0.078	667.400	-0.403	0.687
IC composite:Frequency difference	0.042	0.084	675.500	0.501	0.616
Random effects	Variance	*SD*	Correlation		
Subject (Intercept)	<0.001	<0.001			
Frequency difference	<0.001	<0.001			
Item (Intercept)	0.123	0.350			
IC composite	0.001	0.034	1.000		
Residual	1.411	1.188			

Table 40: Multilevel model *z*-scored AUC estimates for monolinguals in related conditions

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	0.173	0.071	34.000	2.421	0.021
IC composite	0.010	0.053	213.300	0.194	0.847
Frequency difference	-0.043	0.074	667.200	-0.576	0.565
IC composite:Frequency difference	0.064	0.079	675.800	0.808	0.419
Random effects	Variance	*SD*	Correlation		
Subject (Intercept)	<0.001	<0.001			
Frequency difference	<0.001	<0.001			
Item (Intercept)	0.095	0.308			
IC composite	0.003	0.059	1.000		
Residual	1.275	1.129			

Table 41: Multilevel model *x*-axis flips (Poisson distribution) estimates for monolinguals in related conditions

Fixed effects	*b*	*SE*	*z*-value	*p*-value
(Intercept)	1.923	0.034	55.940	<0.001
IC composite	-0.036	0.039	-0.920	0.359
Frequency difference	-0.020	0.025	-0.800	0.425
IC composite:Frequency difference	0.009	0.026	0.340	0.731

Random effects	Variance	*SD*	Correlation
Subject (Intercept)	0.033	0.182	
Frequency difference	<0.001	0.009	-1.000
Item (Intercept)	<0.001	0.000	
IC composite	0.002	0.047	

Visual world RQ3 bilinguals

Visual world full multilevel model output for bilinguals in related conditions. IC composite and LexTALE were centered.

Table 42: Multilevel model log transformed RT estimates for all speakers in related conditions

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	7.174	0.038	41.400	186.789	<0.001
IC composite	-0.008	0.029	38.000	-0.260	0.796
Frequency difference	-0.038	0.020	257.300	-1.907	0.058
LexTALE	-0.002	0.003	37.300	-0.673	0.505
IC composite:Frequency difference	-0.012	0.015	375.200	-0.805	0.421
IC composite:LexTALE	-0.002	0.003	36.200	-0.604	0.550
Frequency difference:LexTALE	0.000	0.002	413.900	0.051	0.959
IC composite:Frequency difference:LexTALE	0.001	0.001	498.400	0.722	0.471
Random effects	**Variance**	***SD***	**Correlation**		
Subject (Intercept)	0.020	0.142			
Frequency difference	<0.001	<0.001	-1.000		
Item (Intercept)	0.004	0.066			
IC composite	0.001	0.035	0.560		
LexTALE	<0.001	0.003	0.130	0.510	
IC composite:LexTALE	<0.001	0.002	0.800	0.560	0.660
Residual	0.032	0.178			

Table 43: Multilevel model *z*-scored AUC estimates for all speakers in related conditions

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	0.015	0.078	49.500	0.197	0.844
IC composite	-0.027	0.052	434.000	-0.531	0.596
Frequency difference	-0.213	0.111	430.400	-1.915	0.056
LexTALE	-0.015	0.006	191.500	-2.329	0.021
IC composite:Frequency difference	-0.109	0.083	479.000	-1.313	0.190
IC composite:LexTALE	0.006	0.006	32.600	0.974	0.338
Frequency difference:LexTALE	-0.014	0.010	487.100	-1.410	0.159
IC composite:Frequency difference:LexTALE	-0.011	0.008	549.500	-1.379	0.168
Random effects	Variance	*SD*	Correlation		
Subject (Intercept)	<0.001	<0.001			
Frequency difference	<0.001	<0.001	-0.440		
Item (Intercept)	0.049	0.221			
IC composite	<0.001	0.019	-0.220		
LexTALE	<0.001	0.008	-0.990	0.350	
IC composite:LexTALE	<0.001	0.016	-0.320	0.990	0.440
Residual	1.126	1.061			

Table 44: Multilevel model *z*-scored MD estimates for all speakers in related conditions

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	0.015	0.078	49.500	0.197	0.844
IC composite	-0.008	0.053	53.000	-0.153	0.879
Frequency difference	-0.185	0.114	53.570	-1.615	0.112
LexTALE	-0.012	0.006	43.120	-1.883	0.067
IC composite:Frequency difference	-0.101	0.085	51.010	-1.182	0.243
IC composite:LexTALE	0.005	0.006	29.860	0.895	0.378
Frequency difference:LexTALE	-0.013	0.010	67.140	-1.276	0.206
IC composite:Frequency difference:LexTALE	-0.012	0.008	50.460	-1.501	0.140
Random effects	**Variance**	***SD***	**Correlation**		
Subject (Intercept)	<0.001	0.008			
Frequency difference	<0.001	0.022	-1.000		
Item (Intercept)	0.080	0.283			
IC composite	0.002	0.048	0.480		
LexTALE	<0.001	0.006	-0.140	0.800	
IC composite:LexTALE	<0.001	0.013	-0.290	0.610	0.890
Residual	1.149	1.072			

Table 45: Multilevel model *x*-axis flips (Poisson distribution) estimates for bilinguals in related conditions

Fixed effects	*b*	*SE*	*z*-value	*p*-value
(Intercept)	1.948	0.063	31.065	<0.001
IC composite	0.048	0.049	0.966	0.334
Frequency difference	-0.030	0.040	-0.748	0.455
LexTALE	0.002	0.006	0.425	0.671
IC composite:Frequency difference	-0.010	0.031	-0.313	0.754
IC composite:LexTALE	0.005	0.005	1.024	0.306
Frequency difference:LexTALE	-0.001	0.004	-0.352	0.725
IC composite:Frequency difference:LexTALE	0.006	0.003	1.788	0.074
Random effects	**Variance**	***SD***	**Correlation**	
Subject (Intercept)	0.056	0.237		
Frequency difference	0.003	0.056		
Item (Intercept)	0.000	0.000		

Phonological priming RQ1

Phonological priming full multilevel model output for all speakers comparing conditions (unrelated, related).

Table 46: Sum coding applied to condition model coefficients for multilevel model log transformed RT estimates

Condition	Contrast codes
Unrelated	1
Related	-1

Table 47: Multilevel model log transformed RT estimates for all speakers

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	373.521	16.111	67.770	23.180	<0.001
condition1	-2.712	3.665	80.020	-0.740	0.461
Random effects	Variance	*SD*	Correlation		
Subject (Intercept)	6153.000	78.440			
condition1	149.000	12.210	-0.570		
Item (Intercept)	6943.000	83.330			
Residual	30318.000	174.120			

Phonological priming RQ3

Phonological priming full multilevel model output for all speakers. The first model includes all target frequencies and the second only higher target frequencies. Target frequency was centered.

Table 48: Sum coding applied to model coefficients for phonological priming multilevel model log transformed RT estimates

Condition	Contrast codes
unrelated	1
related	-1
Speaker type	
monolingual	1
bilingual	-1

Table 49: Multilevel model log transformed RT estimates for all speakers and all target frequencies

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	5.778	0.048	54.790	121.274	<0.001
cond1	-0.004	0.010	117.430	-0.401	0.689
Target frequency	-0.076	0.090	38.430	-0.845	0.403
spkr1	-0.119	0.023	81.520	-5.151	0.000
cond1:Target frequency	-0.003	0.021	113.960	-0.149	0.882
cond1:spkr1	0.006	0.009	111.150	0.627	0.532
Target frequency:spkr1	0.034	0.023	40.160	1.461	0.152
cond1:Target frequency:spkr1	0.054	0.020	91.510	2.672	0.009
Random effects	**Variance**	***SD***	**Correlation**		
Subject (Intercept)	0.035	0.187			
Target frequency	0.002	0.042	-0.220		
cond1	0.001	0.026	-0.450	0.960	
Target frequency:cond1	0.001	0.034	-0.170	0.870	0.760
Item (Intercept)	0.071	0.266			
spkr1	0.001	0.036	0.050		
Residual	0.212	0.461			

Table 50: Multilevel model log transformed RT estimates for all speakers and only higher target frequencies

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	5.755	0.065	22.600	88.801	<0.001
cond1	-0.012	0.014	560.800	-0.822	0.412
IC composite	-0.043	0.028	53.300	-1.535	0.131
spkr1	-0.104	0.027	65.100	-3.927	0.000
cond1:IC composite	0.012	0.014	536.700	0.840	0.401
cond1:spkr1	0.032	0.014	548.200	2.268	0.024
IC composite:spkr1	-0.014	0.026	67.500	-0.541	0.590
cond1:IC composite:spkr1	0.020	0.014	531.400	1.401	0.162
Random effects	Variance	*SD*	Correlation		
Subject (Intercept)	0.032	0.180			
cond1	<0.001	0.022	-1.000		
Item (Intercept)	0.068	0.260			
spkr1	0.001	0.035	0.210		
IC composite	0.002	0.049	-0.050	0.940	
spkr1:IC composite	0.001	0.024	-0.010	0.970	
Residual	0.207	0.455			0.950

Table 51: Multilevel model log transformed RT estimates for monolinguals and all target frequencies

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	5.681	0.053	61.250	107.133	<0.001
cond1	-0.007	0.014	49.220	-0.494	0.624
Target frequency	-0.050	0.093	39.130	-0.541	0.591
IC composite	-0.057	0.035	40.510	-1.617	0.114
cond1:Target frequency	0.048	0.034	49.920	1.405	0.166
cond1:IC composite	0.020	0.016	47.260	1.276	0.208
Target frequency:IC composite	0.029	0.033	163.190	0.875	0.383
cond1:Target frequency:IC composite	0.015	0.037	43.740	0.418	0.678
Random effects	Variance	*SD*	Correlation		
Subject (Intercept)	0.030	0.173			
cond1	0.001	0.029	-0.690		
Target frequency	0.001	0.027	0.020	0.680	
cond1: Target frequency	0.008	0.087	-0.280	0.620	0.770
Item (Intercept)	0.071	0.267			
IC composite	<0.001	0.003	1.000		
Residual	0.220	0.469			

Phonological priming RQ3 bilinguals

Phonological priming full multilevel model output for bilinguals only. The first model includes all target frequencies, the second only higher target frequencies, and the third only lower target frequencies. Target frequency and LexTALE were centered.

Table 52: Multilevel model log transformed RT estimates for bilinguals and all target frequencies

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	5.834	0.062	63.100	94.184	<0.001
cond1	0.001	0.018	65.500	0.064	0.949
LexTALE	-0.009	0.004	37.500	-2.237	0.031
Target frequency	-0.096	0.101	38.300	-0.957	0.345
cond1:LexTALE	0.001	0.002	65.400	0.636	0.527
cond1:Target frequency	-0.073	0.039	649.400	-1.901	0.058
LexTALE:Target frequency	0.003	0.004	78.800	0.742	0.461
cond1:LexTALE:Target frequency	-0.002	0.004	639.300	-0.501	0.616
Random effects	**Variance**	***SD***	**Correlation**		
Subject (Intercept)	0.035	0.186			
cond1	<0.001	0.021	0.360		
Target frequency	0.001	0.037	-0.070	0.910	
cond1:Target frequency	<0.001	0.019	0.980	0.170	-0.270
Item (Intercept)	0.079	0.281			
LexTALE	<0.001	0.002	1.000		
Residual	0.202	0.450			

Table 53: Multilevel model log transformed RT estimates for bilinguals and higher-frequency targets

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	5.784	0.078	33.060	74.131	<0.001
cond1	-0.051	0.027	276.000	-1.888	0.060
LexTALE	-0.009	0.005	35.450	-1.947	0.060
IC composite	-0.056	0.042	34.280	-1.351	0.185
cond1:LexTALE	-0.001	0.002	264.330	-0.380	0.704
cond1:IC composite	-0.016	0.021	250.750	-0.768	0.443
LexTALE:IC composite	-0.002	0.004	34.990	-0.581	0.565
cond1:LexTALE:IC composite	-0.001	0.002	254.680	-0.706	0.481
Random effects	**Variance**	***SD***	**Correlation**		
Subject (Intercept)	0.033	0.182			
cond1	0.000	0.020	1.000		
Item (Intercept)	0.065	0.254			
LexTALE	<0.001	0.000	-1.000		
IC composite	0.002	0.041	-0.090	0.090	
LexTALE:IC composite	<0.001	0.001	0.840	-0.840	0.460
Residual	0.178	0.422			

Table 54: Multilevel model log transformed RT estimates for bilinguals and lower-frequency targets

Fixed effects	*b*	*SE*	*df*	*t*-value	*p*-value
(Intercept)	5.847	0.082	28.800	71.101	<0.001
cond1	0.035	0.026	453.900	1.325	0.186
LexTALE	-0.011	0.005	37.500	-2.386	0.022
cond1:LexTALE	0.003	0.002	453.100	1.042	0.298
Random effects	**Variance**	***SD***	**Correlation**		
Subject (Intercept)	0.034	0.185			
cond1	<0.001	0.010	-1.000		
Item (Intercept)	0.095	0.308			
LexTALE	<0.001	0.004	1.000		
Residual	0.227	0.477			

Appendix G: Proportional proximity *t*-test results

Table 55: Proportional proximity to target compared between conditions (related, unrelated) averaged across speaker type.

Timestamp	Related mean	Unrelated mean	*p*-value	Holm corrected *p*-value
0	0.357	0.354	0.109	1.000
1	0.358	0.356	0.151	1.000
2	0.360	0.357	0.095	1.000
3	0.362	0.359	0.063	1.000
4	0.365	0.360	0.043*	1.000
5	0.367	0.362	0.032*	1.000
6	0.369	0.364	0.03*	1.000
7	0.371	0.365	0.028*	1.000
8	0.372	0.367	0.029*	1.000
9	0.374	0.368	0.029*	1.000
10	0.376	0.369	0.032*	1.000
11	0.377	0.371	0.039*	1.000
12	0.378	0.372	0.04*	1.000
13	0.379	0.373	0.038*	1.000
14	0.381	0.374	0.035*	1.000
15	0.382	0.375	0.032*	1.000
16	0.383	0.376	0.027*	1.000
17	0.384	0.377	0.027*	1.000
18	0.385	0.378	0.027*	1.000
19	0.387	0.379	0.027*	1.000
20	0.388	0.380	0.027*	1.000
21	0.389	0.381	0.026*	1.000
22	0.390	0.382	0.024*	1.000
23	0.392	0.383	0.021*	1.000
24	0.393	0.384	0.021*	1.000
25	0.395	0.386	0.023*	1.000
26	0.396	0.387	0.027*	1.000
27	0.398	0.389	0.035*	1.000
28	0.399	0.391	0.041*	1.000
29	0.401	0.393	0.052	1.000
30	0.403	0.395	0.057	1.000
31	0.405	0.397	0.062	1.000
32	0.407	0.399	0.071	1.000
33	0.409	0.401	0.075	1.000
34	0.411	0.403	0.086	1.000
35	0.413	0.406	0.117	1.000
36	0.415	0.408	0.162	1.000

Timestamp	Related mean	Unrelated mean	*p*-value	Holm corrected *p*-value
37	0.418	0.412	0.210	1.000
38	0.421	0.415	0.235	1.000
39	0.425	0.418	0.241	1.000
40	0.429	0.423	0.245	1.000
41	0.433	0.427	0.262	1.000
42	0.438	0.432	0.306	1.000
43	0.443	0.438	0.391	1.000
44	0.448	0.444	0.491	1.000
45	0.454	0.451	0.596	1.000
46	0.461	0.459	0.707	1.000
47	0.469	0.467	0.786	1.000
48	0.478	0.477	0.875	1.000
49	0.487	0.487	0.965	1.000
50	0.498	0.498	0.940	1.000
51	0.509	0.511	0.773	1.000
52	0.520	0.524	0.584	1.000
53	0.533	0.539	0.390	1.000
54	0.546	0.555	0.228	1.000
55	0.560	0.572	0.122	1.000
56	0.574	0.589	0.069	1.000
57	0.590	0.606	0.041*	1.000
58	0.607	0.624	0.029*	1.000
59	0.624	0.642	0.021*	1.000
60	0.642	0.661	0.014*	1.000
61	0.660	0.680	0.01*	0.924
62	0.679	0.700	0.006*	0.540
63	0.697	0.719	0.003*	0.278
64	0.715	0.738	0.002*	0.152
65	0.733	0.757	0.001*	0.085
66	0.751	0.774	<0.001*	0.047*
67	0.768	0.791	<0.001*	0.029*
68	0.785	0.807	<0.001*	0.035*
69	0.802	0.822	0.001*	0.059
70	0.818	0.836	0.001*	0.085
71	0.833	0.849	0.002*	0.191
72	0.847	0.861	0.003*	0.318
73	0.860	0.871	0.009*	0.788
74	0.873	0.881	0.025*	1.000
75	0.883	0.890	0.041*	1.000
76	0.893	0.899	0.055	1.000
77	0.901	0.906	0.119	1.000
78	0.909	0.912	0.196	1.000

Timestamp	Related mean	Unrelated mean	*p*-value	Holm corrected *p*-value
79	0.915	0.918	0.209	1.000
80	0.920	0.923	0.230	1.000
81	0.926	0.928	0.314	1.000
82	0.930	0.932	0.367	1.000
83	0.934	0.935	0.382	1.000
84	0.938	0.939	0.473	1.000
85	0.941	0.941	0.624	1.000
86	0.943	0.943	0.751	1.000
87	0.945	0.945	0.836	1.000
88	0.946	0.946	0.862	1.000
89	0.947	0.947	0.750	1.000
90	0.948	0.948	0.666	1.000
91	0.948	0.949	0.574	1.000
92	0.948	0.949	0.559	1.000
93	0.948	0.949	0.544	1.000
94	0.948	0.949	0.545	1.000
95	0.949	0.949	0.544	1.000
96	0.949	0.949	0.543	1.000
97	0.949	0.949	0.539	1.000
98	0.949	0.949	0.536	1.000
99	0.949	0.949	0.531	1.000
100	0.949	0.949	0.526	1.000

Table 56: Proportional proximity to competitor compared between conditions (related, unrelated) averaged across speaker type.

Timestamp	Related mean	Unrelated mean	*p*-value	Holm corrected *p*-value
0	0.344	0.343	0.574	1.000
1	0.345	0.345	0.733	1.000
2	0.347	0.346	0.68	1.000
3	0.348	0.347	0.658	1.000
4	0.349	0.349	0.846	1.000
5	0.351	0.351	0.977	1.000
6	0.352	0.352	0.972	1.000
7	0.353	0.354	0.945	1.000
8	0.355	0.355	0.938	1.000
9	0.356	0.356	0.94	1.000
10	0.357	0.357	0.941	1.000
11	0.358	0.358	0.987	1.000
12	0.359	0.360	0.94	1.000
13	0.361	0.361	0.86	1.000
14	0.362	0.362	0.814	1.000
15	0.363	0.364	0.795	1.000
16	0.364	0.366	0.721	1.000
17	0.366	0.367	0.702	1.000
18	0.367	0.368	0.707	1.000
19	0.368	0.370	0.715	1.000
20	0.370	0.371	0.729	1.000
21	0.371	0.373	0.735	1.000
22	0.373	0.374	0.791	1.000
23	0.375	0.376	0.908	1.000
24	0.377	0.377	0.971	1.000
25	0.379	0.379	0.883	1.000
26	0.382	0.380	0.733	1.000
27	0.385	0.382	0.545	1.000
28	0.388	0.383	0.367	1.000
29	0.391	0.385	0.268	1.000
30	0.394	0.387	0.181	1.000
31	0.397	0.389	0.122	1.000
32	0.401	0.391	0.071	1.000
33	0.404	0.394	0.043*	1.000
34	0.409	0.396	0.02*	1.000
35	0.413	0.399	0.01*	0.630
36	0.418	0.402	0.005*	0.351
37	0.422	0.405	0.002*	0.161
38	0.427	0.407	0.001*	0.063

Timestamp	Related mean	Unrelated mean	*p*-value	Holm corrected *p*-value
39	0.431	0.409	<0.001*	0.031*
40	0.434	0.412	<0.001*	0.015*
41	0.439	0.414	<0.001*	0.007*
42	0.443	0.417	<0.001*	0.003*
43	0.447	0.419	<0.001*	0.001*
44	0.452	0.422	<0.001*	0.001*
45	0.456	0.425	<0.001*	<0.001*
46	0.460	0.428	<0.001*	<0.001*
47	0.464	0.430	<0.001*	<0.001*
48	0.467	0.432	<0.001*	<0.001*
49	0.470	0.435	<0.001*	<0.001*
50	0.472	0.436	<0.001*	<0.001*
51	0.473	0.438	<0.001*	<0.001*
52	0.474	0.438	<0.001*	<0.001*
53	0.474	0.438	<0.001*	<0.001*
54	0.475	0.437	<0.001*	<0.001*
55	0.474	0.436	<0.001*	<0.001*
56	0.472	0.434	<0.001*	<0.001*
57	0.469	0.432	<0.001*	<0.001*
58	0.465	0.430	<0.001*	<0.001*
59	0.460	0.427	<0.001*	<0.001*
60	0.454	0.423	<0.001*	<0.001*
61	0.448	0.418	<0.001*	0.001*
62	0.441	0.413	<0.001*	0.003*
63	0.433	0.407	<0.001*	0.006*
64	0.425	0.400	<0.001*	0.01*
65	0.417	0.393	<0.001*	0.013*
66	0.409	0.386	<0.001*	0.015*
67	0.401	0.379	<0.001*	0.021*
68	0.392	0.372	<0.001*	0.034*
69	0.383	0.366	0.001*	0.076
70	0.375	0.359	0.003*	0.211
71	0.366	0.354	0.01*	0.663
72	0.358	0.348	0.031*	1.000
73	0.350	0.343	0.088	1.000
74	0.343	0.338	0.223	1.000
75	0.336	0.333	0.387	1.000
76	0.330	0.328	0.561	1.000
77	0.325	0.324	0.762	1.000
78	0.321	0.321	0.901	1.000
79	0.318	0.318	0.91	1.000
80	0.316	0.315	0.935	1.000

Timestamp	Related mean	Unrelated mean	*p*-value	Holm corrected *p*-value
81	0.313	0.314	0.913	1.000
82	0.311	0.312	0.741	1.000
83	0.310	0.311	0.589	1.000
84	0.308	0.310	0.395	1.000
85	0.307	0.309	0.259	1.000
86	0.305	0.308	0.185	1.000
87	0.305	0.307	0.15	1.000
88	0.304	0.307	0.135	1.000
89	0.304	0.306	0.127	1.000
90	0.304	0.306	0.121	1.000
91	0.303	0.306	0.12	1.000
92	0.303	0.306	0.119	1.000
93	0.303	0.306	0.116	1.000
94	0.303	0.306	0.116	1.000
95	0.303	0.306	0.115	1.000
96	0.303	0.306	0.114	1.000
97	0.303	0.306	0.114	1.000
98	0.303	0.306	0.114	1.000
99	0.303	0.306	0.114	1.000
100	0.303	0.306	0.116	1.000

Table 57: Monolingual proportional proximity to target compared between relative competitor frequencies.

Timestamp	HF competitor mean	LF competitor mean	*p*-value	Holm corrected *p*-value
0	0.351	0.362	0.007*	1.000
1	0.352	0.363	0.008*	1.000
2	0.352	0.364	0.009*	1.000
3	0.355	0.365	0.023*	1.000
4	0.357	0.366	0.075	1.000
5	0.359	0.367	0.109	1.000
6	0.361	0.367	0.212	1.000
7	0.363	0.368	0.361	1.000
8	0.365	0.369	0.522	1.000
9	0.366	0.369	0.663	1.000
10	0.368	0.369	0.78	1.000
11	0.369	0.370	0.876	1.000
12	0.370	0.370	0.985	1.000
13	0.371	0.371	0.958	1.000
14	0.372	0.371	0.922	1.000
15	0.373	0.372	0.892	1.000
16	0.374	0.373	0.869	1.000
17	0.375	0.373	0.822	1.000
18	0.376	0.374	0.786	1.000
19	0.377	0.375	0.775	1.000
20	0.378	0.376	0.762	1.000
21	0.379	0.377	0.777	1.000
22	0.380	0.378	0.812	1.000
23	0.382	0.380	0.802	1.000
24	0.383	0.381	0.759	1.000
25	0.385	0.382	0.731	1.000
26	0.386	0.383	0.689	1.000
27	0.388	0.384	0.628	1.000
28	0.390	0.385	0.528	1.000
29	0.392	0.387	0.468	1.000
30	0.395	0.388	0.415	1.000
31	0.397	0.390	0.391	1.000
32	0.398	0.391	0.397	1.000
33	0.400	0.394	0.434	1.000
34	0.402	0.396	0.472	1.000
35	0.404	0.398	0.51	1.000
36	0.405	0.401	0.595	1.000
37	0.408	0.404	0.659	1.000
38	0.411	0.408	0.739	1.000

Timestamp	HF competitor mean	LF competitor mean	*p*-value	Holm corrected *p*-value
39	0.414	0.411	0.758	1.000
40	0.419	0.416	0.768	1.000
41	0.424	0.421	0.779	1.000
42	0.429	0.426	0.756	1.000
43	0.435	0.432	0.784	1.000
44	0.441	0.439	0.85	1.000
45	0.448	0.447	0.922	1.000
46	0.456	0.456	0.956	1.000
47	0.464	0.467	0.817	1.000
48	0.473	0.478	0.68	1.000
49	0.484	0.492	0.558	1.000
50	0.496	0.505	0.497	1.000
51	0.509	0.519	0.434	1.000
52	0.523	0.534	0.419	1.000
53	0.538	0.549	0.441	1.000
54	0.554	0.565	0.443	1.000
55	0.569	0.582	0.389	1.000
56	0.586	0.599	0.38	1.000
57	0.603	0.617	0.387	1.000
58	0.622	0.635	0.366	1.000
59	0.640	0.655	0.327	1.000
60	0.659	0.674	0.323	1.000
61	0.678	0.691	0.375	1.000
62	0.699	0.709	0.46	1.000
63	0.719	0.727	0.603	1.000
64	0.740	0.744	0.738	1.000
65	0.759	0.762	0.802	1.000
66	0.777	0.780	0.787	1.000
67	0.794	0.799	0.617	1.000
68	0.811	0.817	0.511	1.000
69	0.827	0.834	0.452	1.000
70	0.841	0.849	0.354	1.000
71	0.855	0.862	0.306	1.000
72	0.867	0.874	0.299	1.000
73	0.878	0.884	0.278	1.000
74	0.887	0.893	0.231	1.000
75	0.895	0.901	0.221	1.000
76	0.903	0.908	0.235	1.000
77	0.910	0.915	0.251	1.000
78	0.916	0.920	0.34	1.000
79	0.922	0.924	0.516	1.000
80	0.927	0.928	0.707	1.000

Timestamp	HF competitor mean	LF competitor mean	*p*-value	Holm corrected *p*-value
81	0.931	0.931	0.954	1.000
82	0.935	0.935	0.929	1.000
83	0.938	0.938	0.993	1.000
84	0.941	0.941	0.752	1.000
85	0.943	0.944	0.509	1.000
86	0.945	0.946	0.411	1.000
87	0.946	0.948	0.416	1.000
88	0.947	0.948	0.478	1.000
89	0.948	0.949	0.532	1.000
90	0.948	0.950	0.578	1.000
91	0.949	0.950	0.622	1.000
92	0.949	0.950	0.658	1.000
93	0.949	0.950	0.682	1.000
94	0.949	0.950	0.69	1.000
95	0.949	0.950	0.697	1.000
96	0.949	0.950	0.699	1.000
97	0.949	0.950	0.701	1.000
98	0.949	0.950	0.703	1.000
99	0.949	0.950	0.699	1.000
100	0.949	0.950	0.693	1.000

Note. HF refers to higher-frequency, and LF refers to lower-frequency.

Table 58: Bilingual proportional proximity to target compared between relative competitor frequencies.

Timestamp	HF competitor mean	LF competitor mean	*p*-value	Holm corrected *p*-value
0	0.356	0.358	0.617	1.000
1	0.360	0.360	0.852	1.000
2	0.362	0.362	0.986	1.000
3	0.366	0.365	0.879	1.000
4	0.369	0.368	0.733	1.000
5	0.372	0.370	0.703	1.000
6	0.375	0.373	0.698	1.000
7	0.378	0.375	0.666	1.000
8	0.381	0.377	0.576	1.000
9	0.384	0.379	0.528	1.000
10	0.386	0.382	0.566	1.000
11	0.387	0.384	0.655	1.000
12	0.389	0.387	0.8	1.000
13	0.390	0.389	0.922	1.000
14	0.391	0.391	0.957	1.000
15	0.392	0.393	0.869	1.000
16	0.394	0.395	0.864	1.000
17	0.396	0.396	0.995	1.000
18	0.398	0.397	0.904	1.000
19	0.399	0.398	0.835	1.000
20	0.401	0.399	0.809	1.000
21	0.403	0.400	0.806	1.000
22	0.404	0.402	0.885	1.000
23	0.404	0.404	0.988	1.000
24	0.405	0.407	0.86	1.000
25	0.406	0.409	0.765	1.000
26	0.407	0.411	0.671	1.000
27	0.408	0.413	0.623	1.000
28	0.410	0.415	0.605	1.000
29	0.412	0.417	0.576	1.000
30	0.414	0.419	0.578	1.000
31	0.416	0.421	0.639	1.000
32	0.419	0.422	0.735	1.000
33	0.422	0.424	0.86	1.000
34	0.425	0.425	0.964	1.000
35	0.427	0.427	0.972	1.000

Timestamp	HF competitor mean	LF competitor mean	*p*-value	Holm corrected *p*-value
36	0.430	0.429	0.916	1.000
37	0.432	0.431	0.978	1.000
38	0.434	0.434	0.98	1.000
39	0.438	0.438	0.995	1.000
40	0.441	0.442	0.895	1.000
41	0.444	0.447	0.793	1.000
42	0.446	0.452	0.691	1.000
43	0.449	0.456	0.617	1.000
44	0.452	0.461	0.531	1.000
45	0.456	0.467	0.433	1.000
46	0.460	0.473	0.343	1.000
47	0.465	0.481	0.273	1.000
48	0.471	0.489	0.242	1.000
49	0.477	0.496	0.211	1.000
50	0.484	0.505	0.168	1.000
51	0.490	0.514	0.138	1.000
52	0.498	0.524	0.122	1.000
53	0.507	0.534	0.107	1.000
54	0.516	0.545	0.09	1.000
55	0.525	0.557	0.058	1.000
56	0.536	0.572	0.034*	1.000
57	0.547	0.586	0.022*	1.000
58	0.561	0.602	0.017*	1.000
59	0.578	0.617	0.021*	1.000
60	0.595	0.632	0.03*	1.000
61	0.614	0.649	0.045*	1.000
62	0.633	0.666	0.057	1.000
63	0.652	0.683	0.068	1.000
64	0.670	0.700	0.076	1.000
65	0.688	0.717	0.082	1.000
66	0.706	0.733	0.091	1.000
67	0.723	0.748	0.108	1.000
68	0.742	0.764	0.123	1.000
69	0.759	0.782	0.11	1.000
70	0.776	0.800	0.074	1.000
71	0.793	0.816	0.069	1.000
72	0.810	0.831	0.081	1.000
73	0.828	0.847	0.087	1.000
74	0.844	0.860	0.117	1.000

Timestamp	HF competitor mean	LF competitor mean	*p*-value	Holm corrected *p*-value
75	0.860	0.872	0.175	1.000
76	0.873	0.882	0.266	1.000
77	0.886	0.891	0.479	1.000
78	0.897	0.899	0.798	1.000
79	0.906	0.906	0.951	1.000
80	0.913	0.912	0.828	1.000
81	0.920	0.918	0.766	1.000
82	0.925	0.924	0.785	1.000
83	0.929	0.929	0.989	1.000
84	0.933	0.934	0.789	1.000
85	0.937	0.938	0.584	1.000
86	0.939	0.942	0.348	1.000
87	0.941	0.944	0.318	1.000
88	0.943	0.946	0.364	1.000
89	0.944	0.946	0.389	1.000
90	0.946	0.947	0.469	1.000
91	0.946	0.948	0.418	1.000
92	0.946	0.948	0.41	1.000
93	0.946	0.948	0.412	1.000
94	0.946	0.948	0.405	1.000
95	0.946	0.948	0.401	1.000
96	0.946	0.948	0.399	1.000
97	0.946	0.948	0.4	1.000
98	0.946	0.948	0.399	1.000
99	0.946	0.948	0.396	1.000
100	0.946	0.948	0.399	1.000

Note. HF refers to higher-frequency, and LF refers to lower-frequency.

Table 59: Monolingual proportional proximity to competitor compared between relative competitor frequencies.

Timestamp	HF competitor mean	LF competitor mean	*p*-value	Holm corrected *p*-value
0	0.344	0.347	0.462	1.000
1	0.344	0.348	0.383	1.000
2	0.344	0.349	0.294	1.000
3	0.345	0.351	0.285	1.000
4	0.346	0.351	0.244	1.000
5	0.347	0.353	0.243	1.000
6	0.348	0.354	0.241	1.000
7	0.349	0.355	0.223	1.000
8	0.349	0.357	0.199	1.000
9	0.350	0.358	0.173	1.000
10	0.351	0.360	0.137	1.000
11	0.351	0.361	0.105	1.000
12	0.352	0.363	0.089	1.000
13	0.353	0.364	0.079	1.000
14	0.353	0.365	0.066	1.000
15	0.354	0.366	0.060	1.000
16	0.355	0.367	0.057	1.000
17	0.355	0.368	0.058	1.000
18	0.356	0.369	0.056	1.000
19	0.357	0.370	0.050	1.000
20	0.358	0.371	0.049*	1.000
21	0.358	0.373	0.045*	1.000
22	0.359	0.374	0.043*	1.000
23	0.361	0.375	0.047*	1.000
24	0.362	0.377	0.055	1.000
25	0.364	0.378	0.069	1.000
26	0.367	0.380	0.096	1.000
27	0.370	0.382	0.119	1.000
28	0.373	0.385	0.134	1.000
29	0.376	0.388	0.157	1.000
30	0.380	0.391	0.182	1.000
31	0.383	0.395	0.218	1.000
32	0.388	0.399	0.249	1.000
33	0.392	0.403	0.259	1.000
34	0.397	0.408	0.298	1.000

Timestamp	HF competitor mean	LF competitor mean	*p*-value	Holm corrected *p*-value
35	0.402	0.412	0.327	1.000
36	0.407	0.416	0.404	1.000
37	0.412	0.421	0.451	1.000
38	0.417	0.425	0.515	1.000
39	0.422	0.428	0.600	1.000
40	0.427	0.432	0.693	1.000
41	0.432	0.436	0.744	1.000
42	0.436	0.440	0.760	1.000
43	0.441	0.446	0.708	1.000
44	0.446	0.451	0.681	1.000
45	0.451	0.457	0.650	1.000
46	0.455	0.462	0.623	1.000
47	0.459	0.466	0.631	1.000
48	0.463	0.469	0.644	1.000
49	0.466	0.472	0.662	1.000
50	0.468	0.474	0.694	1.000
51	0.470	0.475	0.728	1.000
52	0.471	0.475	0.781	1.000
53	0.471	0.474	0.836	1.000
54	0.470	0.472	0.879	1.000
55	0.468	0.469	0.939	1.000
56	0.466	0.466	0.988	1.000
57	0.462	0.461	0.952	1.000
58	0.457	0.455	0.911	1.000
59	0.452	0.449	0.824	1.000
60	0.446	0.442	0.760	1.000
61	0.439	0.434	0.707	1.000
62	0.431	0.427	0.731	1.000
63	0.422	0.419	0.825	1.000
64	0.413	0.411	0.904	1.000
65	0.403	0.403	0.960	1.000
66	0.394	0.394	0.993	1.000
67	0.385	0.383	0.893	1.000
68	0.377	0.373	0.724	1.000
69	0.368	0.362	0.583	1.000
70	0.359	0.352	0.442	1.000
71	0.351	0.343	0.361	1.000
72	0.343	0.335	0.331	1.000

Timestamp	HF competitor mean	LF competitor mean	*p*-value	Holm corrected *p*-value
73	0.336	0.329	0.315	1.000
74	0.330	0.323	0.272	1.000
75	0.325	0.318	0.256	1.000
76	0.320	0.314	0.252	1.000
77	0.316	0.310	0.277	1.000
78	0.313	0.308	0.355	1.000
79	0.310	0.307	0.481	1.000
80	0.308	0.305	0.618	1.000
81	0.306	0.304	0.682	1.000
82	0.305	0.303	0.640	1.000
83	0.304	0.302	0.602	1.000
84	0.304	0.302	0.543	1.000
85	0.303	0.301	0.497	1.000
86	0.303	0.301	0.454	1.000
87	0.303	0.301	0.430	1.000
88	0.303	0.301	0.435	1.000
89	0.303	0.301	0.454	1.000
90	0.303	0.301	0.463	1.000
91	0.303	0.301	0.477	1.000
92	0.303	0.301	0.482	1.000
93	0.303	0.301	0.484	1.000
94	0.303	0.301	0.482	1.000
95	0.303	0.301	0.481	1.000
96	0.303	0.301	0.482	1.000
97	0.303	0.301	0.482	1.000
98	0.303	0.301	0.483	1.000
99	0.303	0.301	0.483	1.000
100	0.303	0.301	0.477	1.000

Note. HF refers to higher-frequency, and LF refers to lower-frequency.

Table 60: Bilingual proportional proximity to competitor compared between relative competitor frequencies.

Timestamp	HF competitor mean	LF competitor mean	*p*-value	Holm corrected *p*-value
0	0.343	0.341	0.318	1.000
1	0.346	0.343	0.435	1.000
2	0.348	0.345	0.463	1.000
3	0.350	0.347	0.472	1.000
4	0.352	0.349	0.474	1.000
5	0.354	0.351	0.552	1.000
6	0.355	0.352	0.625	1.000
7	0.356	0.354	0.667	1.000
8	0.358	0.355	0.711	1.000
9	0.359	0.357	0.746	1.000
10	0.360	0.358	0.738	1.000
11	0.362	0.359	0.642	1.000
12	0.364	0.360	0.523	1.000
13	0.366	0.360	0.447	1.000
14	0.368	0.361	0.413	1.000
15	0.370	0.363	0.357	1.000
16	0.372	0.364	0.328	1.000
17	0.374	0.366	0.328	1.000
18	0.376	0.368	0.339	1.000
19	0.378	0.370	0.320	1.000
20	0.381	0.372	0.292	1.000
21	0.383	0.374	0.285	1.000
22	0.386	0.376	0.295	1.000
23	0.388	0.379	0.332	1.000
24	0.390	0.382	0.362	1.000
25	0.392	0.385	0.418	1.000
26	0.395	0.388	0.502	1.000
27	0.397	0.391	0.536	1.000
28	0.400	0.395	0.630	1.000
29	0.403	0.398	0.683	1.000
30	0.405	0.401	0.724	1.000
31	0.408	0.405	0.806	1.000
32	0.410	0.408	0.887	1.000
33	0.412	0.412	0.967	1.000
34	0.416	0.418	0.876	1.000
35	0.419	0.423	0.794	1.000

Timestamp	HF competitor mean	LF competitor mean	*p*-value	Holm corrected *p*-value
36	0.423	0.427	0.731	1.000
37	0.427	0.431	0.728	1.000
38	0.431	0.436	0.728	1.000
39	0.435	0.439	0.767	1.000
40	0.439	0.442	0.815	1.000
41	0.443	0.445	0.866	1.000
42	0.448	0.449	0.957	1.000
43	0.452	0.452	0.968	1.000
44	0.457	0.455	0.903	1.000
45	0.461	0.458	0.885	1.000
46	0.464	0.461	0.858	1.000
47	0.466	0.464	0.885	1.000
48	0.469	0.467	0.893	1.000
49	0.472	0.469	0.865	1.000
50	0.474	0.470	0.823	1.000
51	0.476	0.472	0.790	1.000
52	0.479	0.473	0.723	1.000
53	0.481	0.473	0.615	1.000
54	0.483	0.474	0.558	1.000
55	0.485	0.474	0.483	1.000
56	0.487	0.472	0.357	1.000
57	0.487	0.469	0.269	1.000
58	0.486	0.465	0.198	1.000
59	0.484	0.460	0.145	1.000
60	0.479	0.454	0.118	1.000
61	0.473	0.448	0.109	1.000
62	0.466	0.441	0.105	1.000
63	0.459	0.434	0.097	1.000
64	0.452	0.427	0.096	1.000
65	0.445	0.420	0.092	1.000
66	0.439	0.414	0.090	1.000
67	0.432	0.407	0.093	1.000
68	0.423	0.401	0.103	1.000
69	0.415	0.393	0.101	1.000
70	0.407	0.385	0.089	1.000
71	0.399	0.377	0.065	1.000
72	0.391	0.368	0.058	1.000
73	0.381	0.360	0.056	1.000
74	0.371	0.352	0.072	1.000

Timestamp	HF competitor mean	LF competitor mean	*p*-value	Holm corrected *p*-value
75	0.360	0.345	0.118	1.000
76	0.351	0.339	0.192	1.000
77	0.342	0.335	0.373	1.000
78	0.335	0.332	0.665	1.000
79	0.330	0.329	0.927	1.000
80	0.325	0.326	0.894	1.000
81	0.321	0.323	0.752	1.000
82	0.318	0.321	0.680	1.000
83	0.316	0.318	0.705	1.000
84	0.313	0.315	0.763	1.000
85	0.311	0.312	0.824	1.000
86	0.309	0.310	0.861	1.000
87	0.307	0.308	0.831	1.000
88	0.306	0.307	0.791	1.000
89	0.305	0.307	0.764	1.000
90	0.305	0.306	0.708	1.000
91	0.304	0.306	0.684	1.000
92	0.304	0.306	0.682	1.000
93	0.304	0.305	0.691	1.000
94	0.304	0.305	0.715	1.000
95	0.304	0.305	0.726	1.000
96	0.304	0.305	0.731	1.000
97	0.304	0.305	0.733	1.000
98	0.304	0.305	0.735	1.000
99	0.304	0.305	0.736	1.000
100	0.304	0.305	0.740	1.000

Note. HF refers to higher-frequency, and LF refers to lower-frequency.

Bibliography

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., … Costa, A. (2012). Bilingualism Tunes the Anterior Cingulate Cortex for Conflict Monitoring. *Cerebral Cortex*, *22*(9), 2076–2086. https://doi.org/10.1093/cercor/bhr287

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.

Ammons, R. B. (1955). Rotary pursuit apparatus: I. Survey of variables. *Psychological Bulletin*, *52*(1), 69–76. http://dx.doi.org/10.1037/h0040849

Baayen, H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 309–412. https://doi.org/doi:10.1016/j.jml.2007.12.005

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database*. Philadelphia: University of Pennsylvania, Linguistic Data Corsortium. Retrieved from https://catalog.ldc.upenn.edu/ldc96l14

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459.

Bartolotti, J., & Marian, V. (2012). Language Learning and Control in Monolinguals and Bilinguals. *Cognitive Science*, *36*(6), 1129–1147. https://doi.org/10.1111/j.1551-6709.2012.01243.x

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2016). lme4: Linear mixed-effects models using Eigen and S4 (Version 1.1-12). Retrieved from http://cran.r-project.org/web/packages/lme4/index.html

Baum, S., & Titone, D. (2014). Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics*, *35*(05), 857–894. https://doi.org/10.1017/S0142716414000174

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research*. Timonium, MD: York Press.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development*, *70*(3), 636–644.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12*(01), 3. https://doi.org/10.1017/S1366728908003477

Bialystok, E. (2010). Global–local and trail-making tasks by monolingual and bilingual children: Beyond inhibition. *Developmental Psychology*, *46*(1), 93–105. https://doi.org/10.1037/a0015466

Bialystok, E., Craik, F. I. M., & Luk, G. (2008a). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics*, *21*(6), 522–538. https://doi.org/10.1016/j.jneuroling.2007.07.001

Bialystok, E., Craik, F., & Luk, G. (2008b). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 859–873. https://doi.org/10.1037/0278-7393.34.4.859

Bialystok, E., & Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism: Language and Cognition*, *15*(02), 397–401. https://doi.org/10.1017/S136672891100040X

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(04), 525–531. https://doi.org/10.1017/S1366728909990423

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, *9*(1), 103–119.

Blumenfeld, H. K., & Marian, V. (2007a). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes*, *22*(5), 633–660. https://doi.org/10.1080/01690960601000746

Blumenfeld, H. K., & Marian, V. (2007b). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes*, *22*(5), 633–660. https://doi.org/10.1080/01690960601000746

Blumenfeld, H. K., & Marian, V. (2011). Bilingualism influences inhibitory control in auditory comprehension. *Cognition*, *118*(2), 245–257. https://doi.org/10.1016/j.cognition.2010.10.012

Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in Stimulus–Stimulus inhibition. *Bilingualism: Language and Cognition*, *17*(03), 610–629. https://doi.org/10.1017/S1366728913000564

Blumenfeld, H. K., Schroeder, S. R., Bobb, S. C., Freeman, M. R., & Marian, V. (2016). Auditory word recognition across the lifespan: Links between linguistic and nonlinguistic inhibitory control in bilinguals and monolinguals. *Linguistic Approaches to Bilingualism*, *6*(1–2), 119–146. https://doi.org/10.1075/lab.14030.blu

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America*, *106*(4), 2074–2085.

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes*, *27*(7–8), 1205–1224. https://doi.org/10.1080/01690965.2012.660170

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System*, *36*(1), 22–34. https://doi.org/10.1016/j.system.2007.11.003

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *The Quarterly Journal of Experimental Psychology*, *64*(1), 74–95. https://doi.org/10.1080/17470218.2010.499174

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, *3*, 269–322.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evalutation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, Marc, Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, *20*(3), 530–548.

Canseco-Gonzalez, E., Brehm, L., Brick, C. A., Brown-Schmidt, S., Fischer, K., & Wagner, K. (2010). Carpet or Cárcel: The effect of age of acquisition and language mode on bilingual lexical access. *Language and Cognitive Processes*, *25*(5), 669–705. https://doi.org/10.1080/01690960903474912

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition*, *28*(1), 143–153.

Chrabaszcz, A., & Gor, K. (2014). Context Effects in the Processing of Phonolexical Ambiguity in L2: Context Effects in Processing of L2. *Language Learning*, *64*(3), 415–455. https://doi.org/10.1111/lang.12063

Cook, S.V. (2012). *Phonological form in L2 lexical access: Friend or foe* (Unpublished doctoral dissertation). University of Maryland.

Cook, S.V., & Gor, K. (2015a). Lexical access in L2: Representational deficit or processing constraint? *The Mental Lexicon*, *10*(2), 247–270. https://doi.org/10.1075/ml.10.2.04coo

Cook, S.V., & Gor, K. (2015b). Lexical access in L2: Representational deficit or processing constraint? *The Mental Lexicon*, *10*(2), 247–270. https://doi.org/10.1075/ml.10.2.04coo

Cook, S.V., Pandža, N. B., Lancaster, A. K., & Gor, K. (2016). Fuzzy Nonnative Phonolexical Representations Lead to Fuzzy Form-to-Meaning Mappings. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01345

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, *34*(2), 269–284. https://doi.org/10.1016/j.wocn.2005.06.002

Dahan, D., & Gaskell, M. G. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language*, *57*(4), 483–501.

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*(1), 15–28.

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon*, *8*(3), 372–420. https://doi.org/10.1075/ml.8.3.06dar

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive Advantage in Bilingualism An Example of Publication Bias? *Psychological Science*, *26*(1), 99–107.

Declerck, M., & Philipp, A. M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin & Review*, *22*(6), 1630–1645. https://doi.org/10.3758/s13423-015-0836-1

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first-and second-language word recognition: A lexical entrenchment account. *The Quarterly Journal of Experimental Psychology*, *66*(5), 843–863.

Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, *41*, 495–518.

Dufour, S. (2008). Phonological priming in auditory word recognition: When both controlled and automatic processes are responsible for the effects. *Canadian*

Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, *62*(1), 33–41. https://doi.org/10.1037/1196-1961.62.1.33

Dufour, S., & Peereman, R. (2003). Inhibitory priming effects in auditory word recognition: When the target's competitors conflict with the prime word. *Cognition*, *88*(3), B33–B44.

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review*, *15*(4), 850–855. https://doi.org/10.3758/PBR.15.4.850

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*(2), 345–360. https://doi.org/10.1016/j.wocn.2007.11.002

Farmer, T. A., Cargill, S. A., Hindy, N. C., Dale, R., & Spivey, M. J. (2007). Tracking the continuity of language comprehension: Computer mouse trajectories suggest parallel syntactic processing. *Cognitive Science*, *31*(5), 889–909.

Finkbeiner, M., Almeida, J., Janssen, N., & Caramazza, A. (2006). Lexical selection in bilingual speech production does not involve language suppression. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1075–1089. https://doi.org/10.1037/0278-7393.32.5.1075

Flege, E.J. (1995). Second language speech learning. In *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 680.

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241. https://doi.org/10.3758/BRM.42.1.226

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Giezen, M. R., Blumenfeld, H. K., Shook, A., Marian, V., & Emmorey, K. (2015). Parallel language activation and inhibitory control in bimodal bilinguals. *Cognition*, *141*, 9–25. https://doi.org/10.1016/j.cognition.2015.04.009

Goldinger, S., Luce, P. A., & Pisoni, D. . (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28*(5), 501–518.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814. https://doi.org/10.1016/j.jml.2007.07.001

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, *140*(2), 186–209. https://doi.org/10.1037/a0022256

Gor, K. (2018). Phonological priming and the role of phonology in nonnative word reocgnition. *Bilingualism: Language and Cognition*.

Gor, K., & Cook, S. V. (2018). A mare in a pub? Nonnative facilitation in phonological priming. *Second Language Research*, 18.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(02), 67. https://doi.org/10.1017/S1366728998000133

Hamburger, M., & Slowiaczek, L. M. (1996). Phonological priming reflects lexical competition. *Psychonomic Bulletin & Review*, *3*(4), 520–525.

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, *18*(3), 384–401.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *The Journal of the Acoustical Society of America*, *117*(2), 896–907. https://doi.org/10.1121/1.1823291

Incera, S., & McLennan, C. T. (2016). The time course of within and between-language interference in bilinguals. *International Journal of Bilingualism*, 1–12.

Ito, T. A., Friedman, N. P., Bartholow, B. D., Correll, J., Loersch, C., Altamirano, L. J., & Miyake, A. (2015). Toward a Comprehensive Understanding of Executive Cognitive Function in Implicit Racial Bias. *Journal of Personality and Social Psychology*, *108*(2), 187–218. https://doi.org/10.1037/a0038557.supp

Ju, M., & Luce, P. A. (2004). Falling on sensitive ears constraints on bilingual lexical activation. *Psychological Science*, *15*(5), 314–318.

Kapa, L. L., & Colombo, J. (2014). Executive function predicts artificial language learning. *Journal of Memory and Language*, *76*, 237–252. https://doi.org/10.1016/j.jml.2014.07.004

Kemper, S., Schmalzried, R., Herman, R., Leedahl, S., & Mohankumar, D. (2009). The Effects of Aging and Dual Task Demands on Language Production. *Aging, Neuropsychology, and Cognition*, *16*(3), 241–259. https://doi.org/10.1080/13825580802438868

Kieslich, P. J., Wulff, D. U., Henninger, F., & Haslbeck, J. M. B. (2017). Mousetrap: An R package for processing and analyzing mouse-tracking data (Version 3.0.0). Retrieved from 10.5281/zenodo.290283

Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, *65*(1), 42–73. https://doi.org/10.1016/j.jml.2011.03.002

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 802–823. https://doi.org/10.1037/a0030859

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2018). Tests in Linear Mixed Effects Models (Version 3.0-1). Retrieved from https://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf

La Heij, W. (2005). Selection processes in monolingual and bilingual leical access. In J. F. Kroll & De Groot, A.M.B. (Eds.), *Handbook of bilingualism* (pp. 289–307). New York: Oxford University Press.

Lancaster, A. (2017). *The effects o linguistic properties and inhibitory control on native spoken word recognition*. Unpublished qualifying paper, University of Maryland.

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. https://doi.org/10.3758/s13428-011-0146-0

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 12–31. https://doi.org/10.1037/0278-7393.34.1.12

Lenth, R. (2018). emmeans: Estimated marginal means, aka leas-squares means (Version 1.2.2). Retrieved from https://github.com/rvlenth/emmeans

Lev-Ari, S., & Peperkamp, S. (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics*, *41*(5), 320–331. https://doi.org/10.1016/j.wocn.2013.06.002

Linck, J. A., & Cunnings, I. (2015). The Utility and Application of Mixed-Effects Models in Second Language Research: Mixed-Effects Models. *Language Learning*, *65*(S1), 185–207. https://doi.org/10.1111/lang.12117

Logan, G. D. (1994). On the ability to inhibit thought and action: A user' guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 189–239). San Diego, CA: Academic Press.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*(1), 133–156.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*(2), 202–227. https://doi.org/10.1037/0096-3445.132.2.202

Marian, V., Blumenfeld, H. K., & Boukrina, O. V. (2008). Sensitivity to Phonological Similarity Within and Across Languages. *Journal of Psycholinguistic Research*, *37*(3), 141–170. https://doi.org/10.1007/s10936-007-9064-9

Marian, V., Chabal, S., Bartolotti, J., Bradley, K., & Hernandez, A. E. (2014). Differential recruitment of executive control regions during phonological competition in monolinguals and bilinguals. *Brain and Language*, *139*, 108–117. https://doi.org/10.1016/j.bandl.2014.10.005

Marian, V., & Spivey, M. (2003a). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, *24*(02), 173–193.

Marian, V., & Spivey, M. (2003b). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*(2), 97–115. https://doi.org/10.1017/S1366728903001068

Marslen-Wilson, W.D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63.

Marslen-Wilson, William D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1–2), 71–102. https://doi.org/10.1016/0010-0277(87)90005-9

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

Mercier, J., Pivneva, I., & Titone, D. (2014). Individual differences in inhibitory control relate to bilingual spoken word processing. *Bilingualism: Language and Cognition*, *17*(01), 89–117. https://doi.org/10.1017/S1366728913000084

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. https://doi.org/10.1006/jmla.1998.2602

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Mueller, S. ., & Piper, B. . (2014). The psychology experiment building language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods*, *222*, 250–259. https://doi.org/10.1016/j.jneumeth.2013.10.024

Mueller, S. T. (2012). *The PEBL Pursuit Rotor Task*. Retrieved from http://pebl.sourceforge.net/

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*(4), 273–281.

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition*, *111*(2), 263–269. https://doi.org/10.1016/j.cognition.2008.12.007

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265–278. https://doi.org/10.1016/j.cortex.2015.04.014

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Piper, B. J. (2011). Age, handedness, and sex contribute to fine motor behavior in children. *Journal of Neuroscience Methods*, *195*(1), 88–91. https://doi.org/10.1016/j.jneumeth.2010.11.018

Pivneva, I., Mercier, J., & Titone, D. (2014). Executive control modulates cross-language lexical activation during L2 reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 787–796. https://doi.org/10.1037/a0035583

R Core Team. (2016). R: A language environment for statistical computing (Version 3.3.0). R Foundation for Statistical Computin. Retrieved from http://ww.R-project.org

Radeau, M., Morais, J., & Segui, J. (1995). Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(6), 1297–1311. https://doi.org/10.1037/0096-1523.21.6.1297

Rüschemeyer, S. A., Nojack, A., & Limback, M. (2008). A mouse with a roof? Effects of phonological neighbors on processing of words in sentences in a non-native language. *Brain and Language*, *104*, 132–144.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*(6), 110–114. https://doi.org/10.2307/3002019

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 1–17.

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00137

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, *52*(2), 240–255. https://doi.org/10.1016/j.jml.2004.11.001

Segui, J., & Grainger, J. (1990). Priming word recognition with orthographic neighbors: Effects of relative prime-target frequency. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(1), 65.

Singh, N., & Mishra, R. K. (2015). The modulatory role of second language proficiency on performance monitoring: evidence from a saccadic countermanding task in high and low proficient bilinguals. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01481

Singmann, H. (2018). afex: analysis of factorial experiments (Version 0.21-2). Retrieved from https://github.com/singmann/afex

Slowiaczek, L. M., & Hamburger, M. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(6), 1239.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, *3*, 271–295.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), 10393–10398.

Spivey, M. T., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, *10*(3), 281–284.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., … others. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, *51*(2), 247–250.

Takayanagi, S., Dirks, D. D., & Moshfegh, A. (2002). Lexical and talker effects on word recognition among native and non-native listeners with normal and impaired hearing. *Journal of Speech, Language, and Hearing Research*, *45*(3), 585–597.

Treccani, B., & Mulatti, C. (2015). No matter who, no matter how… and no matter whether the white matter matters. Why theories of bilingual advantage in executive functioning are so difficult to falsify. *Cortex*, *73*, 349–351. https://doi.org/10.1016/j.cortex.2015.07.015

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(3), 616–627. https://doi.org/10.1037//0096-1523.28.3.616

Verbruggen, F., Logan, G. D., & Stevens, M. A. (2008). STOP-IT: Windows executable software for the stop-signal paradigm. *Behavior Research Methods*, *40*(2), 479–483. https://doi.org/10.3758/BRM.40.2.479

Vitevitch, Michael S. (2002). Influence of onset density on spoken-word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 270–278. https://doi.org/10.1037//0096-1523.28.2.270

Vitevitch, Michael S. (2012). What do foreign neighbors say about the mental lexicon? *Bilingualism: Language and Cognition*, *15*(01), 167–172. https://doi.org/10.1017/S1366728911000149

Vitevitch, Michael S., Armbrüster, J., & Chu, S. (2004). Sublexical and Lexical Representations in Speech Production: Effects of Phonotactic Probability and Onset Density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 514–529. https://doi.org/10.1037/0278-7393.30.2.514

Vitevitch, M.S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*(4), 325–329.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *The Journal of the Acoustical Society of America*, *119*(1), 597. https://doi.org/10.1121/1.2141003

Wilcox, R. R., & Keselman, H. J. (2003). Modern Robust Data Analysis Methods: Measures of Central Tendency. *Psychological Methods*, *8*(3), 254–274. https://doi.org/10.1037/1082-989X.8.3.254