Within- and between-language competition in adult second language learners: implications for language proficiency

McCall E Sarrett
Interdisciplinary Graduate Program in Neuroscience
University of Iowa

Christine Shea
Dept. of Spanish and Portuguese
Dept. of Linguistics
University of Iowa

Bob McMurray
Dept. of Psychological and Brain Sciences
Dept. of Communication Sciences and Disorders
Dept. of Linguistics
University of Iowa

Corresponding Author
McCall Sarrett
255N PBSB
Dept. of Psychological and Brain Sciences
University of Iowa
Iowa City, IA 52242
mccall-sarrett@uiowa.edu
319-335-0692

ABSTRACT

Second language (L2) learners must not only acquire L2 knowledge (i.e. vocabulary and grammar), but they must also rapidly access this knowledge. In monolinguals, efficient spoken word recognition is accomplished via lexical competition, by which listeners activate a range of candidates that compete for recognition as the signal unfolds. We examined this in adult L2 learners, investigating lexical competition both amongst words of the L2, and between L2 and native language (L1) words. Adult L2 learners (N=33) in their third semester of college Spanish completed a cross-linguistic Visual World Paradigm task to assess lexical activation, along with proficiency assessment (LexTALE-Esp). L2 learners showed typical incremental processing activating both within-L2 and cross-linguistic competitors, similar to fluent bilinguals. Proficiency correlated with both the speed of activating the target (which prior work links to the developmental progression in L1) and the degree to which competition ultimately resolves (linked to robustness of the lexicon).

Keywords: second language acquisition
spoken word recognition
lexical competition
language proficiency
visual world paradigm

Introduction

In an increasingly globalized world, people must often acquire second and third languages. However, acquiring a new language is not trivial. A great deal of work—both pedagogical and scientific—focuses on how learners acquire the complex *knowledge* that underlies language (phonology, words, and grammar; see Muysken, 2013 for a review). Less recognized is that second language (L2) learners must also be able to rapidly access this new knowledge in real-time, given that the ultimate goal should be the ability to communicate and comprehend ideas and information efficiently in everyday conversation. This is particularly true for the population of adult L2 learners who are acquiring L2 knowledge post-adolescence and primarily in classroom contexts (which we refer to as adult L2 learners here). Despite the crucial importance of these real-time skills for efficient communication, little is known about how these skills develop in adult L2 learners.

Lexical competition. Spoken word recognition is particularly fundamental for L2 learners, because words link sound to syntactic and semantic knowledge, and ultimately to conceptual structures that are not language specific. Recognizing spoken words is no mean feat: Even skilled adults listening to their first language (L1) must carry out a complex cascade of cognitive processes to accurately and efficiently recognize words.

One fundamental challenge to spoken word recognition is time: because words unfold over time, there are early points at which multiple words are possible interpretations of the input. For example, when hearing *chief,* there is a point at which only *chee-* (/tʃi/) has been heard, and when a listener could reasonably expect any number of lexical items (i.e. *cheesy, cheerleader, cheek,* etc). In response, listeners activate a range of lexical items in their mental lexicon as this acoustic information unfolds in time (Marslen-Wilson and Zwitserlood, 1989, Allopenna et al., 1998). These candidates compete for recognition (Marslen-Wilson, 1987, Luce and Pisoni, 1998, Dahan et al., 2001b). As more acoustic information arrives, one of these candidates wins out as the recognized word. This process of lexical competition is critical for efficient and accurate perception of spoken words (Marslen-Wilson, 1987, McClelland and Elman, 1986, Hannagan et al., 2013). For bilingual listeners, lexical competition may be vastly more complicated, as competition may unfold over multiple lexica.

Perhaps the most prominent recent methodology for examining this process is the Visual World Paradigm, which allows researchers to track lexical competition in real time (VWP; Allopenna et al., 1998, Tanenhaus et al., 1995). In a typical VWP task, participants see a set of pictures on a monitor. These pictures represent potential candidates for activation that may be considered (and potentially ruled out) when a spoken word is heard. Participants hear the target word and then select the corresponding picture. While they complete this task, eye movements (Allopenna et al., 1998) or the continuous mouse trajectory (Spivey et al., 2005) are recorded. People generally make fixations to potential goals as part of planning their ultimate motor response. Consequently, the fixations allow us to observe when participants are considering different lexical items as speech unfolds in time.

For example, one might use a picture of a *cheeseburger* (the target word), a *cheerleader* (a cohort, which overlaps in its initial phonemes with the target), and a *grasshopper* (unrelated). In this example, as the participant hears *cheeseburger*, both *cheeseburger* and *cheerleader* may be fixated equally at first, but as more acoustic information comes in, *cheeseburger* will eventually win out and *cheerleader* will be suppressed. The unrelated item, *grasshopper*, acts as a baseline to control for fixations not related to language processing (e.g., to visual search).

Methods like the VWP have revealed the complexity of this competition process in even the L1. These methods have also shown how real time spoken word recognition in the L1 develop slowly in school age children (Rigler et al., 2015) and are related to language outcomes (McMurray et al., 2014, McMurray et al., 2010). Less is known about adult L2 learners. In this group, understanding these competition dynamics may be important for two reasons. First, broader work in bilingualism (reviewed below) has used lexical competition as a domain in which to ask questions about the degree to which the lexica in each language are modular or encapsulated. This has not yet been addressed in the context of adult L2 learners. Second, and more important to the needs of adult L2 learners, resolving competition efficiently (i.e. fully suppressing competitors) may be a crucial skill for becoming a fully proficient L2 user in terms of cognitive fluency (Segalowitz, 2010).

Cross-linguistic lexical competition.

A fundamental question in the psycholinguistics of multilingualism is whether words compete in both languages during recognition. For example, a Spanish-English bilingual hearing the English word *chief* will activate other English competitors such as *cheer, cheek, cheat* (etc.), but is the Spanish *chicle* 'chewing gum' also activated? One possibility is that each language's lexicon is fully encapsulated, and the lexica do not interact during real-time processing. In many situations this may be more efficient, as it effectively halves the number of competitors if the language being spoken is known in advance (e.g., all the words are from English, the speaker is a known English speaker).

Alternatively, listeners may consider words from both lexica. This could arise via several mechanisms: They may not be able to fully suppress a whole lexicon, the two lexica could interact, or there may be just a single lexicon comprising words from both languages. In these cases, when hearing *chief,* listeners may consider Spanish competitors like *chicle.* This raises an additional cognitive challenge as multilinguals must now manage competition among a much larger pool of words. Nevertheless, this kind of flexibility could be useful for multilinguals who regularly code switch between languages, or regularly must communicate in situations where the language of a speaker or of a conversation cannot be known in advance. Activating lexical competitors across multiple languages could lead to more flexible spoken word recognition in these cases.

Several studies have examined lexical competition in multilinguals to address the question of modularity (separate lexica) or interactivity (one encompassing lexica or multiple connected lexica) between languages. Much of this work has utilized a cross-linguistic VWP (see also Preston and Lambert, 1969, Chen and Ho, 1986, Tzelgov et al., 1990 for evidence with a Stroop task). For example, the first such cross-linguistic VWP study found that when Russian-English bilinguals hear a Russian target word (i.e. *fishku,* a gamepiece), they show an increased proportion of looks to an English cohort (*fish*) (Spivey and Marian, 1999). This pattern of cross-language activation has been further shown in bilinguals of many languages: Dutch-English (Weber and Cutler, 2004), Spanish-English (Ju and Luce, 2004), Japanese-English (Cutler et al., 2006), Finnish-French (Veivo et al., 2018), and German-English (Blumenfeld and Marian, 2007). Thus, cross-language competition seems to be a general feature of word recognition in balanced bilinguals across languages.

Much of this work has been primarily concerned with this theoretical question of lexical interactivity in bilinguals—whether or not languages interact during online processing. These studies (Blumenfeld and Marian, 2007, Marian and Spivey, 2003, Ju and Luce, 2004, Veivo et al., 2018, Cutler et al., 2006, Weber and Cutler, 2004) have relied largely on participants who learned their second language early (in childhood or during adolescence), have spent time in immersive language settings, and as result would be considered balanced bilinguals. This leaves open two questions.

First, balanced bilinguals developed this interactive lexical system in the context of language exposure in which both languages are interleaved. We contrast balanced bilinguals, who have acquired their L2 early (pre-adolescence) and concurrently with their L1, and adult L2 learners, who have an already established L1 and are acquiring an L2 later (post-adolescence) and via primarily classroom-based contexts. As we describe, the experience of most adult L2 learners is profoundly different, and there are theoretical reasons that their particular experience may lead to differences in how words are acquired, and consequently how they may be processed. Second, if L2 listeners show cross-linguistic lexical competition, their ability to manage this competition may be important. L2 listeners must be able to efficiently suppress L1 competitors to allow L2 word recognition to proceed. This suggests a key role for L2 proficiency. As we describe, work on spoken word recognition in L1 acquisition offers a useful framing that may help understand the role of proficiency on real-time processes.

Lexical competition and proficiency.

To understand words rapidly, listeners must quickly activate the correct word (the target) and fully suppress competitors. Work on lexical structure, word learning, and L1 acquisition all suggest ways in which L2 exposure and proficiency may shape real-time word recognition.

First, while the dynamics of lexical competition are strongly driven by the unfolding acoustic input (and the temporary ambiguity that creates), the dynamics of this competition are critically related to internal factors such as the ***structure of the lexicon***. These are likely to change as people acquire new words in their second language. One such factor is the density of the competitive environment (e.g., the number of "neighbor" words that compete with a given word). Words with high cohort densities—e.g., *cat*, which shares onset phonemes with many other words like *cap, can, calf,* etc.—are recognized more slowly than words with lower cohort densities (Magnuson et al., 2007). This suggests that the timecourse of how lexical competition unfolds is influenced by the number of active competitors in the mental lexicon, which in turn is likely to be dramatically altered by acquiring a second language. However, density alone is not as good a predictor as frequency-weighted-density (Luce & Pisoni, 1998)—that is, the ability of a competitor to interfere the target is in part a function of how frequently that words occurs, which itself is a proxy for how well learned that word is.

Second, work on ***word learning*** also suggests that newly acquired words—analogous to L2 lexical acquisition—immediately engage in this competition (Magnuson et al., 2003). In particular, work using the subphonemic mismatch paradigm has shown that competition is mediated by inhibitory links between specific words*;* (e.g., neck specifically interferes with net; Dahan et al., 2001b). Recent work has shown that these inhibitory connections are created very rapidly when learning new words, and are present even on the initial day of training (Kapnoula and McMurray, 2016, Kapnoula et al., 2015, and see Lindsay and Gaskell, 2013, Fernandes et al., 2009); thus these same processes are likely to be available in the (more recently learned) L2 lexicon.

Third, the efficiency of lexical competition is related to language proficiency in a variety of ways both as a function of ***L1 acquisition*** and ***individual differences*** in L1 language ability. This appears as two general constructs: activation rate and resolution (Figure 1 for a schematic). First, the overall speed of activating the target (activation rate) and the amount of initial competition appear to develop over time and depend on language learners' experience (Figure 1A). For example, 9-year-old children are slower to fully activate the target than 16 year olds, and they show higher competition at onset (Rigler et al., 2015, see also Sekerina and Brooks, 2007, McMurray et al., 2018). That is, typical development helps listeners extract more information from the signal at each time point.

Second, within the adolescent age group, individual differences in language ability predict the degree to which competitors are fully suppressed late in processing (Figure 1B). Poorer language users are less likely to fully suppress competitors by the end of the lexical competition; that is, they do not fully *resolve* the competition (Dollaghan, 1998, McMurray et al., 2014, McMurray et al., 2019b, McMurray et al., 2010). Similar findings have been observed in reading where activation rate in *written* recognition predicts reading fluency, suggesting once again a link between proficiency and the degree to which competition resolves quickly (Roembke et al., 2019). These studies all suggest that the real-time dynamics of spoken word recognition are broadly shaped by proficiency (both reflected in development as listeners become more skilled), and by individual differences in language ability within an age.

For multilinguals, there is less evidence on the relationship between L2 proficiency and real-time processing. Blumenfeld and Marian (2013) suggests that bilinguals with greater L2 proficiency show greater activation of cross-linguistic competitors from their L2 when listening in their L1, and also suppress those competitors more fully than less proficient bilinguals. The effect of proficiency is similar in written word recognition as well (Veivo et al., 2018, Veivo et al., 2016, see also Qu et al., 2018). This limited body of work has largely considered balanced bilinguals—those who are highly proficient in both languages and have acquired both languages early in life (pre-adolescence). It suggests that language proficiency is associated with general aspects of efficient processing, but as of yet, processing at the early stages of L2 acquisition still requires a more detailed characterization. However, even balanced bilinguals have not yet been characterized within either of their languages along the fundamental dimensions like activation rate or resolution.

There has been no work on late L2 learners in the early stages of L2 proficiency. This leads to an interesting question regarding within-language competition. The L1 work cited above highlights that an important part of becoming proficient in a language is managing *within-language* competition. However, the nature of within-language competition in the L2 remains largely unexplored. This perhaps becomes an even more important issue for adult L2 learners, for whom the L2 lexicon is more fragile. This may be particularly true in the case of L2 classroom learners who acquire the target language outside of a context where it is used on a daily basis. Importantly, the L1 work raises two clear loci in the timecourse of lexical competition (activation rate and resolution) for a potential relationship with L2 proficiency and how this relationship may play out in real-time L2 competition.

The most relevant work on proficiency and lexical competition comes from the aforementioned L1 literature which shows that "adult-like" activation rates develop slowly and take many years to mature (Rigler et al., 2015, McMurray et al., 2018), but that differences in language ability are correlated not so much with early activation but with the ultimate degree of resolution at the end of processing (McMurray et al., 2019b, Rigler et al., 2015). If variation in L2 proficiency reflects variation in the developmental/learning progression (e.g., some L2 learners are further along than others), then results should pattern like the L1 developmental work (see Figure 1A; Rigler et al., 2015, Sekerina and Brooks, 2007, McMurray et al., 2018). In this case,

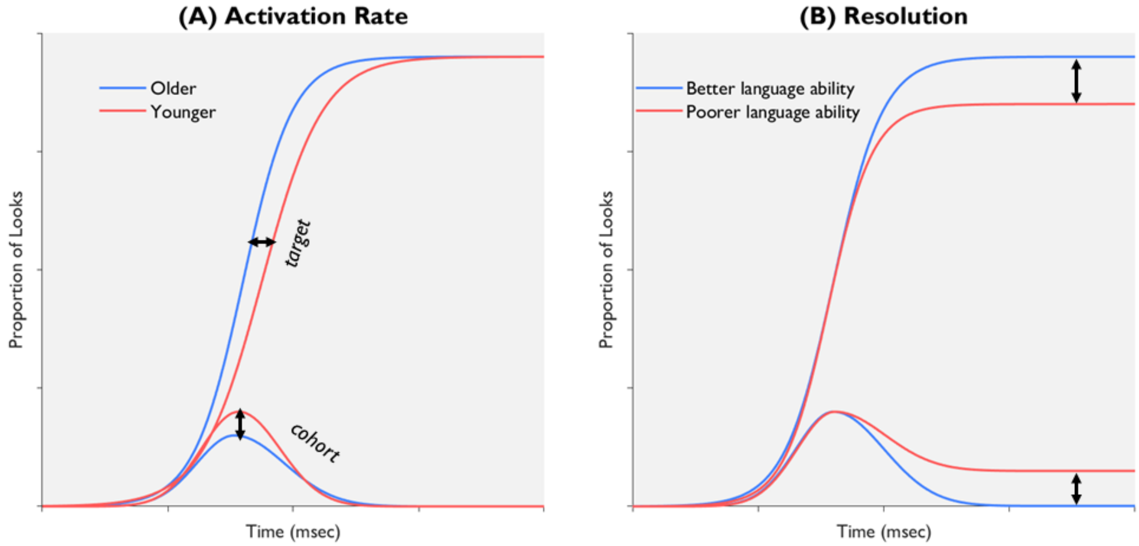

Figure 1. (A) Differences in activation rate due to the timecourse of development from younger to older children appear as delays in target timing and in cohort suppression. (B) Differences in resolution due to individual language ability and robustness of the lexicon emerge later as competition that has not fully resolved—lower maximum asymptote for targets and higher offset baseline for cohorts.

proficiency should be most closely related to the early properties of the fixation curves and to timing based properties of these curves (e.g., the slope of the target, the height of the competitor). That is, more proficient learners will show quicker early activation in their L2. On the other hand, if proficiency in L2 learning reflects not only the exposure of the language but also the robustness of the lexicon and its organization, fixations as a function of L2 proficiency should pattern like the work on individual differences in language ability (see Figure 1B; Dollaghan, 1998, McMurray et al., 2014, McMurray et al., 2019b, McMurray et al., 2010). In this case, L2 proficiency will be most closely related to later properties of the fixation curves (e.g., the asymptotes) with more proficient learners showing more complete resolution of the competition.

Because lexical competition is critical for accurate and efficient speech processing, it is a crucial building block for all subsequent language learning, and also for becoming a skilled L2 listener. But it is unclear how to map the constructs of competition and resolution onto L2 learning because its late onset (relative to the L1) could result in both slower development (predicting differences in activation rates) or less robust language representations (predicting differences in resolution). Thus, by understanding how proficiency is reflected in the timecourse of processing, we can better identify which cognitive differences in word recognition are crucial for language ability in a given population of L2 learners.

The distinct progression of language acquisition in adult L2 learners.

The majority of work in cross-linguistic competition comes from work on balanced bilinguals. It is unclear whether this speaks to adult L2 learners. Furthermore, the ability to manage competition from words in the L1 may be crucial for L2 proficiency—but this could be even more challenging in L2 learners for whom the L1 is highly dominant, and the L2 may be barely present. Adult L2 learners show a distinct trajectory of acquisition from balanced bilinguals: they aren't simply just acquiring the L2 later. Rather, they are starting L2 acquisition from a different place altogether: They already have a large lexicon in their L1 when they begin L2 learning, and this may shape their language development in ways that make them distinct from bilinguals who acquire both languages simultaneously.

So what makes the L2 progression different from L1 and balanced bilingualism? Adult L2 learners have a unique developmental experience that lead to clear differences in their cognitive and neural systems. Childhood balanced bilingual learners differ from adult L2 learners in a number of both productive and perceptual domains—including lexicosemantic processing, overlap of shared neural substrates, grammatical proficiency (Muñoz and Singleton, 2011, Bhatia and Ritchie, 2014). This suggests that L2 listeners may fundamentally differ from balanced bilinguals, but it does not explain why.

The literature on learning and memory offers some insight. This work has shown that when learning two sets of materials, the sequence matters. When materials from both sets are blocked (one set is trained, followed by the other), learning is generally less robust and more fragile than when the items from each set are interleaved (when presentation mimics a "back-and-forth" pattern) (Rohrer, 2012). Further, in blocked presentation, there can be retro-active or pro-active interference. In retro-active interference, information in the second body of material destabilizes learned representations from the earlier one (McCloskey and Cohen, 1989, Spivey and Mirman, 2001), or in pro-active interference, the first set of materials makes it more difficult to learn the second one (i.e., the converse: Jonides and Nee, 2006, Anderson and Neely, 1996). There is also substantial work in procedural learning on the closely related phenomenon of contextual interference (Shea and Morgan, 1979, Magill and Hall, 1990), which again suggests better and more robust learning when materials are interleaved.

The early experience of language acquisition in a balanced bilingual child most closely resembles interleaving. In childhood bilinguals, both languages are learned simultaneously—language-in-use may switch from person to person, or even moment to moment between conversations. This concurrent acquisition may lead to more interconnected cross-linguistic lexica, which would give rise to cross-linguistic competition in these bilinguals in a way that later classroom-based learning may not.

In contrast, classroom-based adult L2 learning (as is the case for the participants studied here) more closely resembles blocking. Adult learners have a previously established native language, and then are exposed to a new language for defined blocks of time in the classroom. This may lead to less stable lexical representations in the L2 (proactive interference). And this in turn could have marked effects on their ability to manage within L2 competition. This blocked learning may also lead to a more encapsulated L2 lexica, or to explicit strategies for suppressing the L1 lexicon, and these in turn could lead to differences in cross-language competition.

These predictions are supported by recent work on the development of inhibitory connections in newly learned L1 words (c.f. McMurray et al., 2016). Models of word recognition suggest that a critical marker of whether a newly learned word has been fully integrated into the lexicon is whether or not newly learned words directly inhibit familiar words (Davis and Gaskell, 2009, Leach and Samuel, 2007). Newly learned words thus must become part of this network, and newly learned words can show these inhibitory effects on known words (Kapnoula et al.,

2015) and amongst themselves (Magnuson et al., 2003) with very little exposure. Crucially, in monolinguals, interleaving facilitates the formation of these inhibitory linkages between newly learned words and previously known words (Lindsay and Gaskell, 2013). However, we have not yet identified the language learning circumstances which may give rise to these linkages (whether classroom-based learning can sufficiently elicit these linkages, or if immersive experiences are necessary) and the timecourse over which they develop (whether interactive linkages between lexica are immediately established interactively or whether these interactions take a longer time to develop).

The Present Study.

The present study had two goals. First, we asked if adult L2 learners showed evidence for both within-language and cross-language lexical competition in their L2. Second, we asked how individual differences in language proficiency correlate with several aspects of real-time spoken word recognition: early activation and later resolution of both withing L2 and cross-language lexical competition.

Participants were adult L2 learners, native English speakers who were enrolled in their third semester of college level Spanish. These courses belong to a four-course general language requirement sequence and are not part of the Spanish major or minor. To measure lexical competition, participants completed a Spanish version of the Visual World Paradigm, similar to that developed by Marian and Spivey (2003). This task assessed both within-L2 (Spanish-Spanish) competition, and between-L2/L1 (Spanish-English) competition. To quantify L2 proficiency, participants took the LexTALE-Esp (Izura et al., 2014), a short assessment designed specifically for use in experimental studies.

With respect to our first question, if cross-language competition (via interconnected lexica or a single encompassing lexicon) is not influenced by age of acquisition or learning context, we would expect to see the same type of between-L2/L1 competition on adult L2 learners as has been observed in early bilinguals. If, however, differences in the language developmental progression between adult L2 learners and balanced bilinguals leads to lexica that are differently organized in adult L2 learners (i.e. modular lexica), then we may not see between-L2/L1 competition.

Our second question is whether individual differences in proficiency and affect speech processing in both L1 and L2. It is not yet known—even for balanced bilinguals—exactly which components of the dynamics of processing are most important for proficiency (activation vs. resolution, and within- vs. between-language competition). We thus estimated different properties of the timecourse of word recognition by fitting curves to the eye-tracking data for each participant, which results in a set of parameters that capture, for example, how quickly participants look to the target item or how fully they suppress a competitor. These were then correlated to general proficiency in the L2.

Methods

Participants. Participants were 34 University of Iowa undergraduates who were enrolled in their third-semester of college Spanish. All participants were native English speakers. Participants with exposure to any language other than English before the age of 2 or who had had an immersive experience with Spanish (e.g. study abroad) were ineligible to participate. Participants completed the study for a small monetary stipend. One participant was excluded due to poor calibration of the eye-tracker. The final sample included 33 participants (6 male, 28 female; approximate age range: 18-24 years old).

Design. We used a cross-linguistic version of the Visual World Paradigm based on Marian and Spivey (2003). Participants saw a set of four pictures on a screen and heard a Spanish word. Participants chose which of the four pictures matched the word they heard by clicking on its referent with a mouse. Words were chosen from the same first- and second-year Spanish textbooks used in the participants' courses, so there was a high likelihood of exposure to and familiarity with the words and their meanings.

Each item set consisted of two phonologically-related pairs: a Spanish-Spanish cohort pair and a Spanish-English cohort pair. Both items in the Spanish-Spanish pair were phonologically and semantically unrelated to the Spanish-English pair. Thus, the two Spanish-English words could serve as unrelated baseline when one of the Spanish-Spanish words was a target (and vice versa). Word length within a pair was controlled as much as possible, such that each item in the pair had roughly the same length. For the Spanish-English pairs, we were concerned about differences in lexical stress across Spanish and English. For nouns, Spanish lexical stress tends to fall on the penultimate syllable when the final syllable is open while English stress is somewhat less regular. Thus, we primarily used two syllable words for these pairs, which both languages stress on the first syllable. If it was not possible to find word pairs of the same syllable length (as we were restricted to using words that early learners would be familiar with), then stress pattern was controlled within a pair such that the location of primary stress was roughly consistent.

Cohorts were created by using pairs that overlapped in the first few phonemes. Spanish-Spanish pairs (e.g., *cielo* [ˈsjelo], 'sky' and *ciencia* [ˈsjensja] 'science') overlapped an average of 2.66 phonemes (range 1-5). Spanish-English pairs overlapped such that a few phonemes of one of the Spanish words (e.g., *botas* [ˈbotas] 'boots' and 'border' [ˈbɔrdər] *frontera*) overlapped with the English name of the competing picture an average of 2.1 phonemes (range: 1-3, average: 2.1). The direct Spanish translation of the English word had no phonemes in common with the Spanish target. As much as possible, given the previous constraints, we controlled for phonemic overlap in the English translation of the Spanish word (i.e. cognates), and for frequency and semantic distinctiveness in each word pair.

This design resulted in three conditions: 1) Spanish target-Spanish competitor; 2) Spanish target-English competitor, and 3) No Competitor condition (see Figure 2). The Spanish-Spanish condition came from the Spanish-Spanish pair, where (from our previous example) either *cielo* or *ciencia* was the target word (and the other was the competitor). In this condition, the Spanish-English pair in the item set (*botas* and *frontera*) function as the unrelated baseline items.

The Spanish-English condition came from the Spanish-English pair, when, for example, *botas* was the target word and *border* (*frontera*) was the competitor. In this condition, the Spanish-Spanish pair (*cielo* and *ciencia*) function as unrelated baseline items. The No-Competitor condition also came from the Spanish-English pair, where *frontera* ('border') was the target word. There is no phonological competitor in this condition, because *frontera* doesn't overlap with the Spanish *botas* (nor with the English translation of *botas*, *boots*, nor with either item from the Spanish-Spanish pair, *cielo* or *ciencia*). The Spanish-Spanish condition had twice as many trials as both the Spanish-English condition and the No-Competitor condition (since in the Spanish-Spanish condition both *cielo* and *ciencia* can function as target and cohort interchangeably).

There were 30 sets of 4 items (Appendix A), and each item was repeated 4 times as the target word (each repetition was a different recording of that word, as described below),

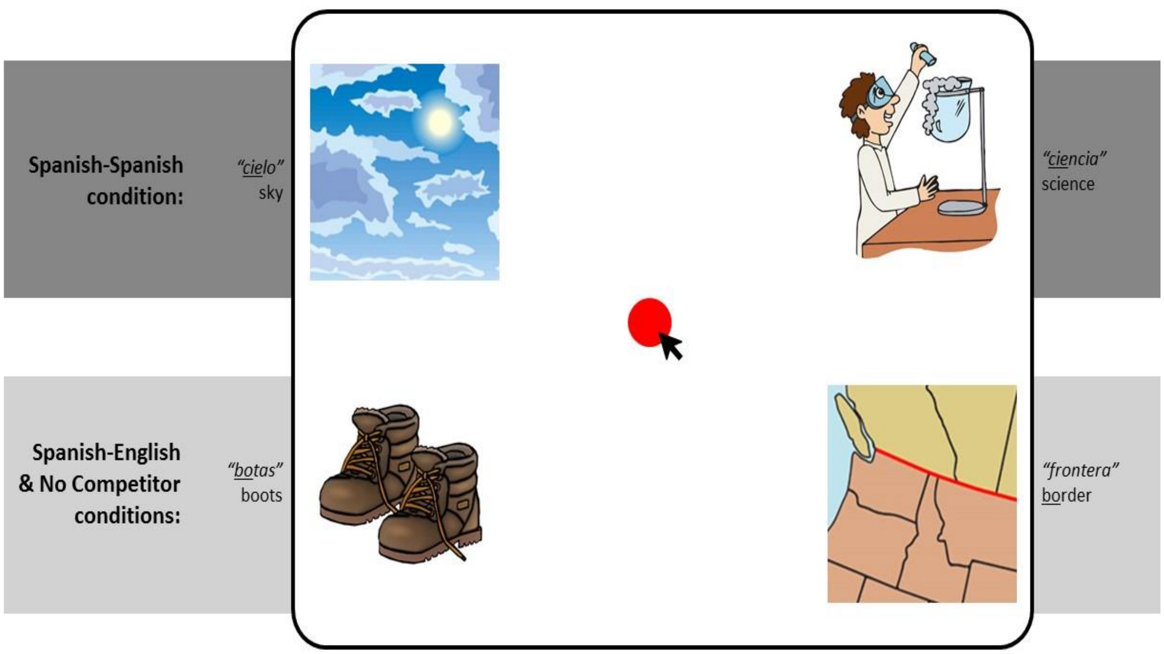

Figure 2. Example screen from the cross-linguistic 4AFC Visual World Paradigm task. Item sets consisted of one Spanish-Spanish competitor pair (top; ex. cielo-ciencia) and one Spanish-English competitor pair (bottom; ex. botas-border). This resulted in three conditions: Spanish-Spanish condition (cielo is the target word, ciencia is the competitor, and vice versa), Spanish-English condition (botas is the target, border *is the cross-linguistic competitor), or No Competitor condition (frontera is the target; botas functions as another unrelated).*

yielding a total of 480 trials for each subject. Each item set consisted of one Spanish-Spanish cohort pair and one Spanish-English cohort pair, such that the pair in the opposite condition functioned as "unrelated" items for some trials. Consequently, there were 240 trials/subject for assessing Spanish-Spanish competition (we could test activation for *cielo* given *ciencia,* and for *ciencia* given *cielo*), however, there were only 120 trials/subject for Spanish-English competition (we could test activation for *frontera* 'border', given *botas,* but when hearing *frontera*, there was no competitor). Items occurred an equal number of times in each of the four corners of the screen, such that position on the screen was not an explicit indicator of which word would be the target or how the items were paired.

Materials. Pictures were developed using a standard lab protocol (McMurray et al., 2010). For each word, we downloaded 8-10 images from a commercial database. These were then viewed by focus groups of native English speakers who selected the best image and identified any modifications that needed to be made. Pictures were then edited (if necessary) to ensure canonical orientations and colors and to remove distracting background elements. Finally, the pictures were approved by a senior lab member who is experienced in the VWP, and each was verified by native Spanish speakers to ensure that the images accurately depicted both the Spanish target word and the English translation. Pictures were matched in style and salience.

Target words were natural utterances from a female speaker who was a native speaker of a Mexican Spanish and a fluent English bilingual. None of the words were produced with the /θ/ characteristic of Peninsular Spanish varieties. Target words were recorded in mono at 44100 Hz, noise reduced, cleaned of clicks and pops, cut at the nearest zero crossings, and normalized to 70 dB using Praat (Boersma, 2006). Four separate recordings of each target word were prepared, so that each repetition of an item was a unique exemplar. One hundred msec of silence was added to the beginning and end of recordings.

Procedures. When participants arrived in the lab, they were greeted in English and all subsequent instructions were given in English. Participants gave informed consent and completed a short language background and demographic questionnaire. Participants then completed the LexTALE-Esp, and the Spanish VWP. The full experimental session took approximately one hour.

LexTALE-Esp. The original Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer and Broersma, 2012) was developed based off of the Eurocentres Vocabulary Size Test (EVST; Meara and Jones, 1987, 1990). It has since been adapted for usage in many other languages, in our case, Spanish (LexTALE-Esp; Izura et al., 2014). In the LexTALE, participants are given a list of words and nonwords, and are instructed to mark only the items they believe to be words in that language. Four of the items in the LexTALE-Esp list of words were also included in the VWP. Two were in the Spanish-Spanish condition (*cuchara* 'spoon', and *cabello* 'hair'), and two were in the Spanish-English condition (*alfombra* 'carpet', and *tiburón* 'shark').

The LexTALE-Esp is scored using a signal detection theory approach. This takes into account the number of words that participants correctly identify, the nonwords they correctly reject, false alarms (a word response to a nonword), and misses (a nonword identified as a word). This measure of vocabulary size has been shown to correlate with individual differences in language processing across a number of different measures (reaction times in a masked neighborhood priming task; Andrews and Hersch, 2010, written word recognition; Chateau and Jared, 2000, word identification times; Diependaele et al., 2013, lexical decision task performance; Yap et al., 2008)

Procedure: Visual World Paradigm. The Spanish VWP was built using SR Research's eBuilder software (SR Research Experiment Builder, 2011). At the beginning of the VWP, participants completed a short familiarization, in which they saw items' pictures and names (orthographically), in Spanish only.

On each trial, four images appeared in the corners of a 1280 x 1024 pixel CRT monitor, 50 pixels away from the edges of the monitor. A small red dot (60-pixel diameter) appeared in the center of the screen, which turned blue after 1000 msec, signaling the start of the trial. During this forced pre-scan period, participants could inspect the four items on the screen to identify the pictures and their locations. Next, participants clicked on the blue dot to play a word over Sennheiser HD 280 Pro over-ear headphones. They then clicked on the picture of the word they heard. This cleared the screen display; there was an intertrial interval of 250 msec before it advanced to the next trial.

Eye-tracking. Eye movements were recorded with a head-mounted EyeLink II eye-tracker from SR Research. Participants were calibrated a standard 9-point calibration, and participants performed a drift correction every 30 trials. Point of gaze was computed from both pupil and corneal reflection at 500 Hz. Data were parsed into blinks, saccades, and fixations automatically by the EyeLink II. For analyses, saccades and the subsequent fixation were collapsed into a single unit, a "look", using EyelinkAnal software (McMurray, 2017). Looks began at the onset of the saccade, lasted through the end of the fixation, and were directed to the mean point of gaze of the fixation. We excluded any look which had an onset prior to 300 msec post-trial onset (this includes 200 msec of assumed oculomotor planning, and 100 msec of silence at word onset), as these were not likely to be driven by the auditory signal . In computing the object each look was referring to, we added 100-pixels to the boundary of the picture. This helped account for imprecise looks, imprecise eyetracking, or other sources of noise in the data. This did not cause any overlap amongst the areas of interest for the pictures.

Results

Accuracy and Reaction Time. We started by examining the accuracy and reaction time of the mouse-click response in the VWP to ensure that subjects were able to complete the task. Participants correctly identified the target words on average of 94.8% (SD = 7.74%). This was not significantly different across conditions (Spanish-Spanish: 95.2%, Spanish-English: 95.3%, No Competitor: 94.2%; $F(2, 96) = 0.218$, $p = .805$). Only trials in which the target word was correctly chosen were included in subsequent analyses of the fixation data. Reaction times averaged

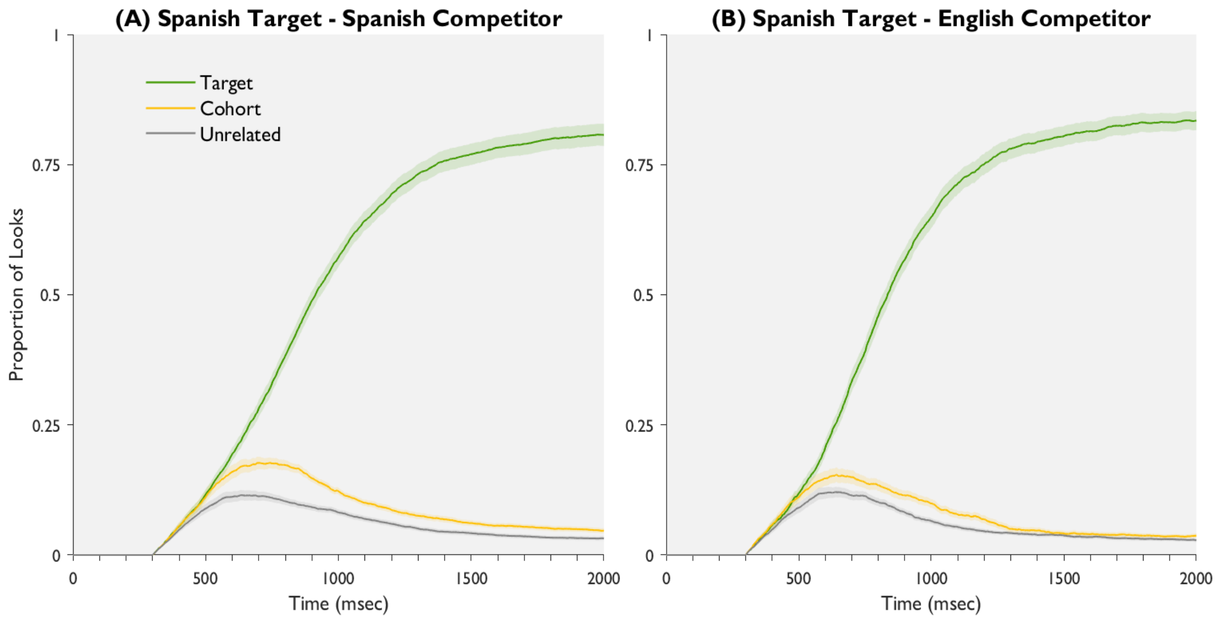

Figure 3. (A) Proportion of looks over time as a function of item type in the within-Spanish condition. Participants showed significantly more peak looks to the Spanish cohort compared to the unrelated. Participants also showed a significantly higher offset baseline in the cohort condition compared to the unrelated. (B) Proportion of looks over time as a function of item type in the cross-linguistic Spanish-English condition. Participants showed significantly more peak looks to the English cohort compared to the unrelated item. Participants also showed a significantly higher offset baseline in the cohort condition compared to the unrelated.

1783.0 msec (SD = 325.5 msec). This also was not significantly different across conditions (Spanish-Spanish RT: 1787.1 msec, Spanish-English: 1713.3 msec, No Competitor: 1848.5; $F(2, 96) = 1.439$, $p = .242$).

Quantifying Lexical Competition. Figure 3A shows the proportions of fixations over time in the Spanish-Spanish condition as function of time and item type (target word, cohort, or unrelated). The unrelated curve is calculated as an average of looks to either unrelated item for that condition. Early on, subjects were equally likely to fixate both the target and the cohort. Cohort looks peaked at around 700 msec (400 msec if you subtract the 200 msec oculomotor delay and the 100 msec of silence at word onset), and then returned to baseline levels. This pattern is remarkably consistent with what one would expect within the L1. Figure 3B shows a similar pattern for the Spanish-English condition. Initially, the Spanish target and the English cohort are equally fixated. Looks to the English cohort peak around 700 msec, and then return to the baseline.

For statistical analyses, we fit non-linear functions to each participants' fixation curves over time. These functions have parameters that capture meaningful properties of the fixations (following McMurray et al., 2010). Target fixations tend to start off slow, ramp up, and reach an asymptote. Thus, target fixations were fit with a logistic function with four free parameters: the minimum and maximum asymptotes, the crossover point (the point on the time axis where looks are halfway between the asymptotes) and slope (the derivative at the crossover). Here, the slope and crossover map to activation rate, and the maximum asymptote to the degree of resolution. In contrast, cohort and unrelated fixations build to a peak, and then fall-off to an asymptotic value. Thus, these were fit using an asymmetric Gaussian function with six free parameters: the onset and offset baseline, the peak height, and timing (the time at which it reaches peak), and the onset and offset slopes. Here, the peak and timing map to activation while the offset slope and baseline correspond to resolution. After fitting the curves, the estimated parameters for each subject can be used in standard statistical analyses as dependent variables which describe specific aspects of the timecourse of fixations. This procedure was repeated averaging across subjects, but grouping by items for item analyses (Clark, 1973).

Functions were fit with a constrained gradient descent method that minimizes the least squared error between the data and the curve while ensuring that the curve remains between 0 and 1, and that the parameters are within reasonable values (e.g., the lower asymptote is below the upper asymptote, the crossover is within the time range) (c.f. McMurray, 2020, version 20). Fits were conducted separately for each subject, each type of referent, and in each condition.

We evaluated the goodness of the fit by computing the correlation between the data and the estimated curve, and by visually comparing the estimated curve to the raw eye-tracking data for each subject. Any curves which did not show adequate fits (by visual inspection or by correlation) were refit by manually re-specifying starting parameters. A poor fit was typically due to the fact that the gradient descent method stopped at a local minimum; in this case, refitting from an alternative starting point was usually sufficient to ensure that the curve reflected the global minimum (best possible fit). Only a few curves had to be refit for each condition and each word type (Spanish-Spanish condition: target (3 refits), cohort (2), unrelated (0); Spanish-English condition: target (4 refits), cohort (0), unrelated (0), No Competitor condition: target (4 refits), unrelated (0).

The resulting fits were very good. In the Spanish-Spanish condition, we obtained overall very good fits for targets (average R=0.99, SD=.001), cohorts (average R=0.98, SD=0.013), and unrelated items (average R=0.99, SD=0.010). Fits were similarly good in the Spanish/English condition for targets (average R=0.99, SD=.001), English cohorts (average R=0.96, SD=.041), and unrelated items (average R=0.97, SD=0.020), and in the No Competitor condition for targets (average R=0.99, SD=0.003) and unrelated items (average R=0.97, SD=0.021). A few subjects' curves had relatively low correlations (3 cohort curves in the Spanish-English condition had $R \leq 0.85$). Visual inspection revealed that they captured the underlying pattern of the data well, and that the low correlations were a result of noise due to fewer trials in that condition. Thus, no subjects were dropped (and these subjects are included in the summaries above).

Within-L2 lexical competition in L2 learners.

We started by asking whether the L2 learners shows the same pattern of lexical competition as is well established for the L1: Do Spanish cohorts participate in lexical competition (Figure 3A)? To do this, we quantified whether the proportion of fixations to the Spanish competitor were greater than that to the unrelated item. For this we used the estimated peak height and asymptotes of the asymmetric gaussian function. Peak height was significantly higher for the Spanish cohorts than the unrelated item ($t_1(32)=10.42$, $p<.0001$; $t_2(29)=6.83$, $p<.0001$). Offset baseline was also significantly higher for the cohorts compared to the unrelated ($t_1(32)=6.93$, $p < .0001$; $t_2(29)=2.57$, $p = .015$). This suggests that L2 learners did not fully suppress the Spanish cohort, even at the end of processing (unlike what would be expected for L1 competition).

Cross-linguistic L2-L1 competition suggests interconnected lexica in adult L2 learners.

Next, we asked whether L2 learners displayed cross language competition: Were English cohorts fixated more than phonologically unrelated items (Figure 3B)? If so, this would suggest that L2 learners show a similar consideration of words in both lexica as in L1; if not, it may indicate that L2 learners have strategies for functionally encapsulating or shutting off their L1 lexicon such that these do not interact in realtime processing during the early stages of language acquisition in adult L2 learners.

In the Spanish-English condition, peak height was again significantly higher for English cohorts than the unrelated item, ($t_1(32)=4.80$, $p<.0001$, $t_2(29)=3.11$, $p=.004$). Furthermore, offset baseline was also significantly higher for English cohorts than unrelated items, by subjects ($t_1(32)=2.32$, $p = .027$) though this was not significant by item ($t_2(29)=.95$, $p=.35$). Again, L2 learners may not fully suppress English competitors, though this may not have been consistent across all items (consistent with the idea that some items may be more robustly learned than others).

An alternative account

One alternative explanation for the cross-linguistic effect seen in the prior analysis is that L2 learners were doing some kind of learning or perceptual grouping throughout the experiment. In this case, the significant looks to the English cohort would have arisen if participants picked up on how the items were paired in the design (e.g., they began to associate *botas* (‘*boots*’) and *frontera* (‘*border*’). This is unlikely given that the unrelated items were also equally likely with the two Spanish-English Target words. However, it is possible that subjects implicitly noticed that the Spanish-Spanish pair was related, and then reasoned by extension the other two Spanish-English items may be as well. This could lead to enhanced English cohort looking, even if lexica are not interconnected.

To address this, we can look at the No Competitor condition. Our design paired a Spanish word with an English cohort—for example *botas* with “border”. However, listeners never heard this English translation throughout the experiment, as all target words were always presented in Spanish. The Spanish translation of "border" is *frontera*, which does not share any phonological overlap with the Spanish *botas*. If the effect were driven purely by associations between items during the experiment, we would expect to see a significant increase in looks to *botas* when *frontera* is the target word.

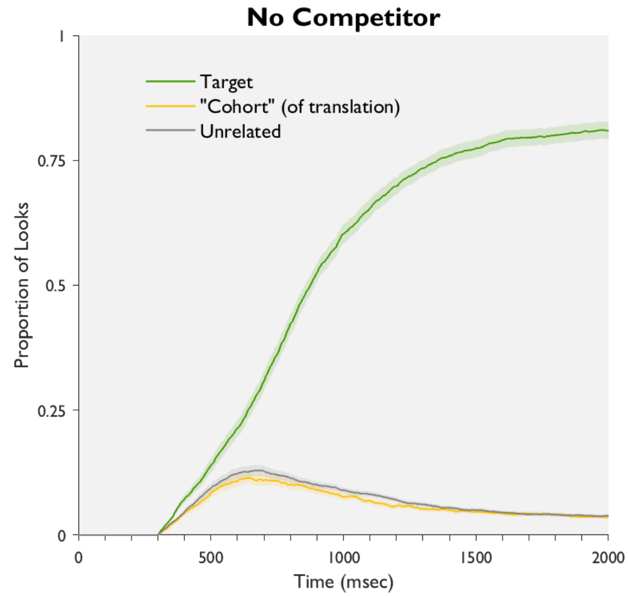

Figure 4. Proportion of looks over time in the no competitor condition. Peak looks to the "cohort" of the translation were not significantly different when compared to the unrelated. Offset baseline also did not differ between the two conditions.

Figure 4 shows the proportion of looks over time in the No Competitor condition. We did not find significant differences between looks to the cohort of the English translation (i.e. *botas*) when participants heard *frontera* (border), and looks to the unrelated item (peak: $t_1(32)=1.09$, $p=.284$; $t_2(29)=0.39$, $p=.70$; offset baseline: $t_1(32)=0.17$, $p=.864$; $t_2(29)=0.42$, $p=.67$). This suggests that our English cohort effect is not due to some sort of associative learning effect over the course of the experiment.

How do within- and between-language competition differ?

Next, we asked whether the magnitude of the competition effects differed for within- and between-language competitors. Figure 3 suggests that the Spanish competitor showed stronger activation (a higher peak) than the English competitor. However, the Spanish competitors also had significantly more phonological overlap with their target words than the English competitors (Spanish-Spanish: M=2.66 phonemes; Spanish-English: M=2.1 phonemes; $t(29) = 2.72$, $p = .01$). This could explain the heightened effects.

To more directly compare the two conditions, we identified a subset of 10 items that controlled for phonological overlap such that all words had only a 2 phoneme overlap. Then, we conducted exploratory analyses to ask whether Spanish competitors behave differently than

English competitors in adult L2 learners. Of course, it was also possible that the baseline (unrelated) fixations differed across conditions. Thus, within each condition, we subtracted the baseline unrelated items from the cohort curves. These difference curves are shown in Figure 5 with the Spanish-Spanish condition in black, and the Spanish-English condition in grey. These difference curves were not guaranteed to take the asymmetric Gaussian shape, and we had no a priori predictions about which aspects of the curves could differ. Thus, we quantified differences between them using the bootstrapped difference of timeseries (BDOTS; Seedorff et al., 2018) which simply identifies time regions over which the curves differed using a novel family-wise error correction.

In this procedure, curves were first fit to the subsetted data following the same procedure previously described. This was done separately for the cohort and unrelated fixations. Again we saw very good fits for the Spanish-Spanish conditions (cohorts: average R=0.97, SD=0.017, unrelateds: average R=0.98, SD=0.012), and moderately good fits for the Spanish-English condition (cohorts: average R=0.89, SD=0.089, unrelateds: average R=0.92, SD=0.078). Only one subject's Spanish-English cohort curve needed to be refit; all others were fit well initially. In the Spanish-English condition, 9 cohort fitted curves and 4 unrelated fitted curves had low correlation values ($R \leq 0.85$). These were inspected visually and were deemed to generally follow the trend of the underlying data; the low value was judged to be due to noise and the fits were retained in the analysis.

Next, these fitted curves underwent a parametric bootstrapping to estimate the standard error of the curves. This was done using the difference of differences option which estimates the cohort-unrelated curve at each point (for each condition) and its associated standard error. After that, a paired t-test is run comparing the two conditions at each time point. Finally, an adjusted alpha is computed using the autocorrelation of the t-statistics ($\rho=0.990$) to account for the fact that that the tests are not fully independent (in this case, $\alpha^* = 0.001$).

Time windows in which the conditions significantly differed are shown with the red bars in Figure 5. This identified four time windows: at 372-654 msec, the cross-language competitor was greater than within; however, at 796-936 msec, 1252-1652 msec, and 1916-2500 msec the converse was true. These exploratory results suggest that (1) English competitors show earlier onset of activation than Spanish competitors, but (2) Spanish competitors are not suppressed as fully as English competitors, when stimuli are matched on other dimensions.

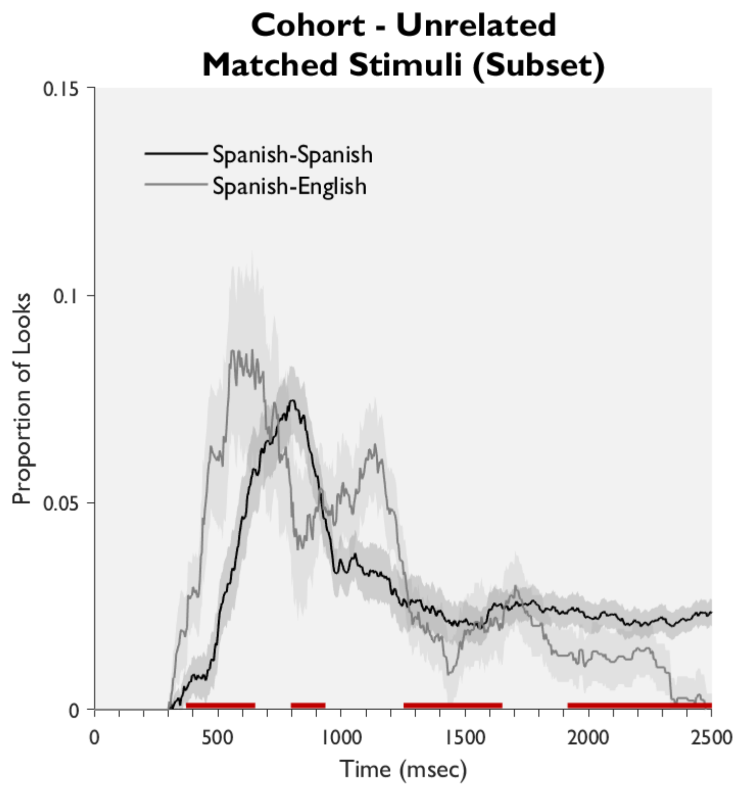

Figure 5. Looks to the unrelated item subtracted from looks to the cohort over time in the Spanish-Spanish condition (black) and the Spanish-English condition (grey). The data has been subsetted such that Spanish and English competitors had the same amount of initial phonological overlap across conditions. Significant time windows are shown by the red bar, calculated using a bootstrapped difference of timeseries (BDOTS).

Individual differences in vocabulary proficiency correlate with online speech processing.

Finally, we asked whether individual differences in vocabulary proficiency correlated with online processing by relating performance on the LexTALE-Esp to the VWP results. Our first analyses focused on the target fixations. We started with a simple data visualization, by conducting a median split on the LexTALE-Esp scores. L2 learners with a LexTALE-Esp score of 7 or below were identified as the low performance group (shown in red in the Figures 6 and 7); those above a 7 were classified in the high performance group (shown in blue in Figures 6 and 7). This suggested robust differences in the target fixations between lower and higher proficiency L2 learners. We quantified this statistically by relating LexTALE proficiency to several key properties of the fixation record. This was done using correlations (not a median-split) to capture the full range of performance.

First, to assess activation rate we combined the crossover and slope parameters of the logistic into a single "timing" parameter (c.f. McMurray et al., 2019a). To calculate timing, we first log-scaled the slope and then computed the Z-score for crossover and slope separately. Next, the crossover Z-score was multiplied by -1 (since larger slopes mean a faster-rising curve, but later crossovers mean a slower rising curve). Finally, the two values were averaged. Second, to assess the degree of resolution, we used the maximum parameter of the targets. Each was computed separately in each of the three conditions: Spanish-Spanish to assess activation rate (timing) and resolution (max) when the task stressed within-language competition; Spanish-English to assess activation rate and resolution when the task stressed cross-language competition; and the no competitor condition when the task stressed primarily non-pictured competitors. Each of these values was separately correlated with performance on the LexTALE-Esp.

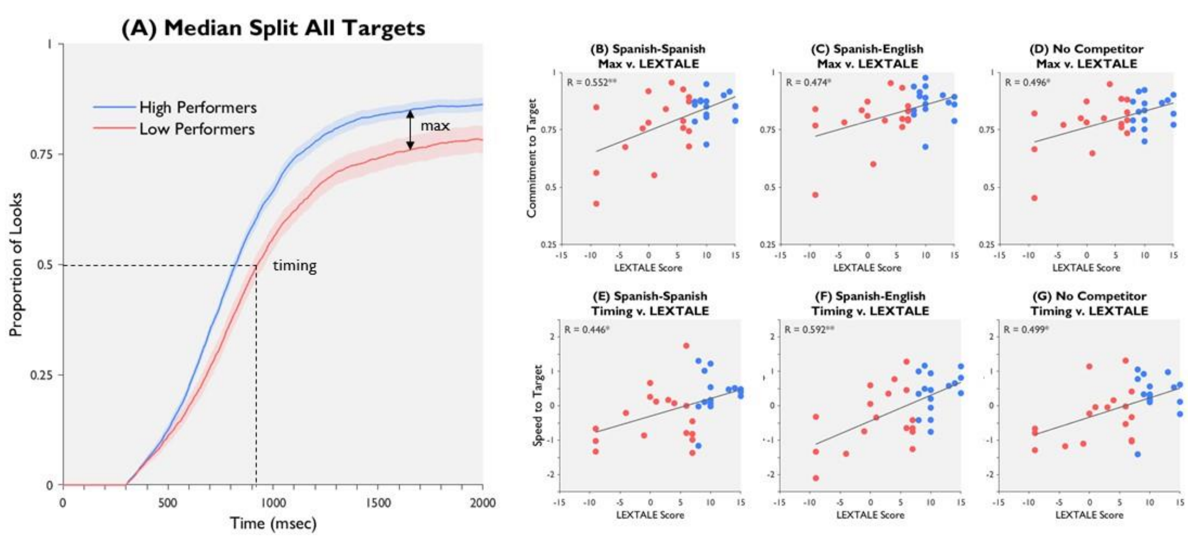

Figure 6. (A) Average looks to the target in all conditions, split by high (blue) and low (red) performers on LexTALE-Esp. The top row of scatterplots shows maximum looks to target in each condition vs. LexTALE-Esp: (B) Spanish-Spanish; (C) Spanish-English; (D) No Competitor. The bottom row of scatter plots shows timing in each condition vs LexTALE-Esp: (E) Spanish-Spanish; (F) Spanish-English; (G) No Competitor.

Figure 6A, shows average looks to target across all conditions, split by high performers (blue) and low performers (red). The top row of scatter plots (Figure 6B, C, and D) show the correlation of LexTALE-Esp with maximum looks to target across different conditions. We found that LexTALE-Esp significantly correlated with maximum asymptote in all three conditions: (B) Spanish-Spanish R=0.552, t(31)=3.689, p<.001; (C) Spanish-English R=0.474, t(31)=3.001, p=.005; (D) No Competitor R=0.496, t(31)=3.177, p=.003. L2 learners who scored better on the vocabulary tests have higher maximum asymptotes.

The bottom row of scatter plots (Figure 6D, F, and G) shows the correlation of LexTALE-Esp with timing. We found that in each condition, LexTALE-Esp significantly correlated with timing of looks to the target word: (E) Spanish-Spanish R=0.446, t(31)=2.777, p=.009; (F) Spanish-English R=0.592, t(31)=4.090, p<.001; (G) No Competitor R=0.499, t(31)=3.209, p=.003. The difference in correlation sizes between the Spanish-Spanish condition and the Spanish-English condition did not differ (z=-1.523, p=.128) (see Lee and Preacher, 2016). L2 learners who had better vocabulary proficiency, as measured by LexTALE-Esp, also were significantly quicker to look to the target word in general.

Both correlations were in the predicted direction and suggest that variation in proficiency is associated with both activation rate and resolution (unlike, for example, the effect of L1 development which is only associated with activation rate). While the Spanish-English correlation for timing was numerically much larger than the others, the difference between the correlations of LexTALE-Esp score with Spanish-Spanish and the Spanish-English condition was not significant (z=1.016, p=.31). This suggests managing competition from L1 and managing within-L2 competition may be similarly important for proficiency.

Next, we turned to the competitors. Figure 7A shows average looks to the cohort across all conditions, split by high performers (blue) and low performers (red) for visualization. The two insets (Figure 7B, C) show significant correlations of LexTALE-Esp score with cohort resolution, measured by the ending baseline, in both the Spanish competitor condition (Figure 7B, left insert: R=-0.395, t(31)=2.396, p=.023) and the English competitor condition (Figure 7C, right insert: R=-0.372, t(31)=2.232, p=.033). These correlations were not significantly different between conditions

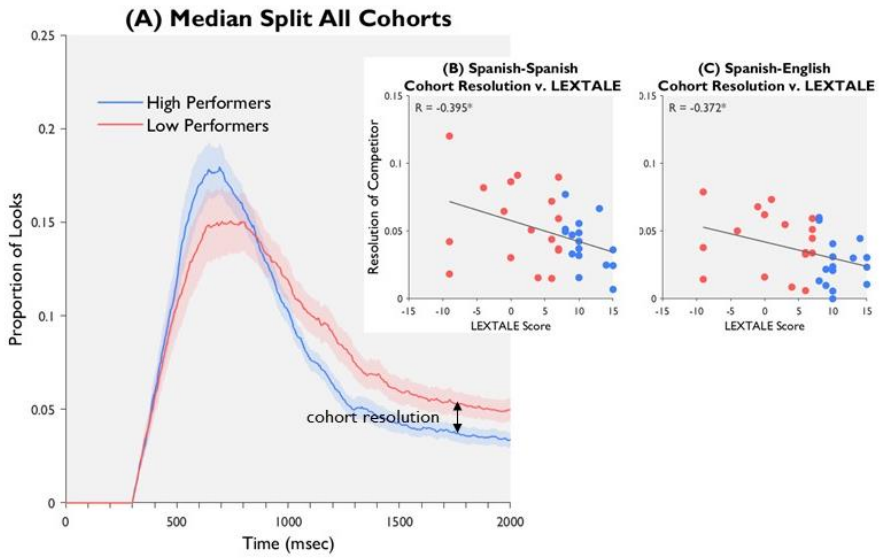

Figure 7. (A) Average looks to the cohort across conditions, split by high (blue) and low (red) performers on LexTALE-Esp. The subplots show correlation of cohort resolution vs. LexTALE-Esp score in (B) the Spanish-Spanish condition and (C) the Spanish-English condition.

(z=-0.187, p=.85). L2 learners with higher vocabulary performance show better ultimate suppression of both the Spanish competitor and the English competitor. Peak cohort fixations (height) did not significantly correlate with LexTALE-Esp in either condition (Spanish-Spanish: R=0.237, t(31)=1.355, p=.18; Spanish-English: R=0.034, t(31)=0.192, p=.85). This could mean that the overall amount of competition (within-L2 or cross-linguistically) is not related to L2 proficiency; it may be likely more of a factor of the unfolding auditory input.

Discussion

This study sought to characterize within-L2 and cross-linguistic lexical competition dynamics in adult L2 learners and to determine if these dynamics are related to L2 language proficiency. First, we asked first whether L2 learners exhibit both between-L2/L1 and within-L2 lexical competition during spoken word recognition. Second, we asked how L2 proficiency is related to several aspects of the timecourse of word recognition. We used a Spanish version of the Visual World Paradigm to measure lexical processing, and the LexTALE-Esp to quantify proficiency.

Within-L2 Competition

We found that even adult L2 learners who are relatively early in their Spanish acquisition show typical incremental processing in their L2 (Figure 3A). This is despite not having much experience with the language (and in the case of our participants, no immersive experience). This fits with the current literature on monolingual word learning in L1, which suggests that new words are immediately integrated into an interconnected lexicon on the same day as they are learned (Magnuson et al., 2003, Kapnoula and McMurray, 2016, Kapnoula et al., 2015). Thus, this may indicate that even when building the lexicon of a new language, a similar rapid establishment of a fully interconnected lexical network is possible.

In general, the overall pattern of lexical competition dynamics in the L2 showed a similar pattern to what is seen in monolingual word recognition. However, there was one notable difference. The offset baseline for looks to the L2 cohort never reached baseline (it remained significantly greater than the unrelated). This suggests that lexical competition may not resolve as fully in adult L2 learners as in monolingual listeners. Functionally, this means that adult L2 learners exhibit greater uncertainty with regards to which lexical item should have been activated, even at these late stages of competition. This may derive from an inability of the lexical system to "complete the deal"—that is, it can generally arrive at the right word, but perhaps some of the contributing mechanisms that normally help this process complete may not be robust enough (specifically for these recently learned L2 words). There are a number of such mechanisms thought to aid more complete resolution of competition which may explain this pattern: lateral inhibition among words (Dahan et al., 2001b), decay (McMurray et al., 2010), mismatch between the input and the competitor (Frauenfelder et al., 2001), and possibly domain-general cognitive control (Zhao et al., 2020, Zhang and Samuel, 2018). It is possible that one or some of these factors contribute to the incomplete resolution seen in adult L2 learners.

It is also possible that full resolution of lexical competition may require more experience with a language—the relatively minimal exposure to Spanish in the classroom may simply not be enough to build the kind of lexical network that supports complete competitor resolution. However, if this is the case, it would appear to conflict with the work on L1 acquisition where even younger children are showing complete resolution (Rigler et al., 2015). Nevertheless, these mechanisms may simply play out differently in L2 acquisition. Perhaps complete resolution of competition is never quite achieved by late L2 learners. That is, the late developmental onset of their language instruction prevents their L2 lexicon from ever developing this capacity. As in children with DLD—who also show this pattern—the complete resolution of competition may only develop as a consequence of "typical" language development—the late onset of L2 acquisition may be a risk factor for this. Future work with more advanced adult L2 learners will be needed to clarify these issues.

Cross-linguistic Competition

We found clear evidence for cross-linguistic competition, even in these relatively early L2 learners (Figure 3B). When hearing a word in Spanish, we observed increased looks to the English cohort, over and above the baseline unrelated object. L2 learners do not seem to have an explicit strategy for suppressing the other lexicon (i.e. their L1, given that the entire VWP task was in Spanish). Our participants were undergraduate adults in their third semester of college Spanish. Importantly, none of these participants had had any sort of immersive experience with Spanish. Therefore, they were acquiring this new language via a truly blocked learning paradigm. In this context, it may have been easier (if not maladaptive) to develop the ability to suppress the other lexicon. Despite this, our participants' L2 and L1 lexica were still functionally interleaved.

We see three possible mechanistic accounts for this. First, one possibility is that word recognition processes are not actually separated by language, that is, there may not be multiple lexica in bilinguals. Instead, there may be one overarching, encompassing lexicon that spans both languages. Our results, and the current literature as a whole, make it difficult to refute this possibility.

Second, we've argued that a critical contrast between adult L2 learners and balanced bilinguals is the difference between blocked and interleaved exposure to the second language. Our results suggest it that blocked learning may not be critical for establishing interconnected lexica in adult L2 learners. This could be because, even in a blocked paradigm, learners are acquiring their new language in a way that is mediated by their L1. In the classroom, L2 learners are often taught to memorize translation equivalents of new L2 vocabulary and are typically given explicit instruction with regards to grammatical structure of the new language. This may result in initial L1-mediated links as the new L2 lexicon is being built, which in turn could generalize and lead to broader and more widespread interconnectedness.

A final possibility is that there are separate lexica, but the entire phenomenon of cross-linguistic competition is driven primarily by temporary ambiguity in the unfolding stimulus. When *chi-* is heard, it is consistent with both *chief* and *chicle* (gum). If the two separate lexica are both engaged with the incoming acoustics at all times, this effect would still be observed, regardless of whether the lexica are not interlinked or not. In this case, the question becomes: why are L2 learners or fluent bilinguals not able to "turn off" an irrelevant lexicon in situations where it is clearly not necessary?

At present, our data cannot rule out any of these accounts, nor are they mutually exclusive. However, it is hard to deny that that bottom-up temporary ambiguity may play at least some role in these findings (and indeed, in the balanced bilingual findings). Thus, it may be fruitful to examine a broader range of competitors (e.g., rhymes or anadromes) which are less susceptible to this issue, or to examine both spoken and written word recognition where the temporal unfolding of the input is less of a factor (c.f. Hendrickson et al., submitted).

Lastly, it is important to point out that while we found evidence for cross-language competition, we also observed that participants here never fully suppressed the L1 competitors (e.g., the offset baseline, see Figure 3B). This effect was weaker than the lack of competitor suppression for within-L2 cohorts, but still significant. Decay and/or cognitive control type mechanisms should already in place for English words. Thus, this points to differences in lateral inhibition amongst words as a potential cause. Spanish targets may have been unable to fully inhibit their English competitors. This could be because some words better integrated than others (especially as this effect was significant by subjects, but not by items). A more robust way to parse out these cross-linguistic links would be to utilize a cross-linguistic subphonemic mismatch paradigm (Dahan et al., 2001b). In this paradigm, for example, fine-grained coarticulatory cues would be used to prime a competitor word in the L1, in order to measure its effects on the L2. This would allow for a more direct measure of cross-linguistic lexical inhibition.

Contrasting Within- and Across-Language Competition

Exploratory analyses asked if there were differences between the two competitor types—within L2 and between L2/L1 (Figure 5). For this analysis, the competitor pairs were matched on the degree of phonological overlap. These are exploratory analyses that relied on only a subset of the items, so they must be treated with caution. However, these analyses found three distinct patterns.

First, English competitors were fixated earlier than Spanish competitors. This may be due to more efficient processing of L1. The participants' native and dominant language was English, and participants have much more experience with that language. Thus, this suggests that at least early in processing, English competitors may dominate—even when the context was clearly Spanish word recognition. This may simply be something analogous to a frequency effect (which also typically appears in the earliest fixations: Dahan et al., 2001a). This would appear to predict that this early L1 bias might be related to proficiency—subjects with less of a bias are more proficient. However, this was not observed: greater fixations to the L1 competitor (cohort height) was not significantly correlated with L2 proficiency. It is possible that cohort height is driven to a large extent by temporary ambiguity in the acoustic signal, and thus it may be less sensitive to proficiency given our relatively small sample size. It is also possible that this correlation was not observed simply because the correlational work relied on all of the items, not just the matched ones. However, to interpret the null correlation seriously: L2 proficiency may just be unrelated to how well L2 learners can suppress initial activation from the irrelevant lexicon. This would make sense if the initial activation of words is obligatory, driven primarily by ambiguous acoustic input as it unfolds. This issue merits further exploration with items that are more targeted to this particular question (e.g., better matched).

Second, it is important to note that we did not find any difference in the peak height between Spanish and English competitors (e.g., Figure 5). This is surprising as the Spanish phonetics in the input should actually favor the Spanish competitor over the English one. This is because the distinct phonology of the Spanish input (e.g., the fact that Spanish uses prevoiced /b/'s while English does not; the fact that /e/ is a monophthong in Spanish but has an offglide in English) will better match the listeners' lexical templates for Spanish words than for English words. This has been shown in balanced bilinguals (Ju and Luce, 2004), but we did not manipulate the acoustics to test this hypothesis (as in Ju and Luce). However, it raises the possibility that adult L2 learners may be less sensitive to the L2 phonology at such a fine-grained level (at least at this relatively early timepoint in language acquisition).

Third, late in the timecourse of processing, English competitors were suppressed more completely than the Spanish competitors. This may result from participants having greater skill in managing L1 competition. That is, their long history with L1, and the fact that it was acquired from birth gives them the capacity to more effectively suppress L1 words—even when listening to their L2. Or it could be due to the greater effort required to activate Spanish words, due to the possibly more fragile nature of their lexical representations. A number of aforementioned factors that contribute to resolution of competition (e.g.. lateral inhibition, decay, cognitive control mechanisms) may also play a role. However, it is unclear the extent to which these mechanisms may be influential across languages.

Individual Differences in Proficiency

Finally, we asked whether individual differences in proficiency are related to the dynamics of within- and across-language competition. We found a strong correlation between LexTALE-Esp score and speed of looking to the target word in all three conditions (Spanish-Spanish, Spanish-English, and No Competitor; Figure 6E, F, & G). This suggests that more proficient L2 learners recognize target words more quickly. It makes sense that activation rate and proficiency would be tightly linked: this patterns with work in L1 development. As monolinguals gain more exposure in their L1, these early word recognition processes speed up (Rigler et al., 2015). This could function similarly for our adult L2 learners: Perhaps the more proficient learners have received more input in their L2—either by studying more often, reading more books, or paying more attention in class.

However, we are unable to exclude that a process of inference may also have played a role in these results. For example, lower proficiency L2 learners are likely less familiar with L2 vocabulary in general. If there is a trial where a participant does not know the target word, but instead must deduce it from which items the word is *not*, their time to finding the target word is going to be delayed (even if they eventually respond correctly). However, our target words were all high frequency words taken from early Spanish textbooks, and accuracy on the VWP task was near ceiling. Furthermore, if this were the case, we would expect looks to the target word to be severely delayed—at least by the time it takes to make one eye movement (~250 msec). However, the effect we observe is much smaller, with about a 100 msec difference between the highest proficiency and lowest proficiency learners. Nevertheless, we can't rule this out as a contributing factor.

Our results suggest that overall L2 proficiency is in part achieved by more rapid building of lexical activation of the target word, regardless of what type of competitor is used as a distractor (whether L2 or L1). An important caveat to this result is that non-displayed competitors can influence speed of target fixations (Magnuson et al., 2007). In our experiment, we controlled which competitors were presented on the screen, but not which competitors were *considered* by the participant. This may make it more difficult to make direct comparisons across conditions for the speed of target activation. However, minimally, we can conclude that managing competition from both languages is essential for becoming proficient in a new L2. It may be that these same dynamics play a role in lexical activation for both languages. This raises an interesting question: To what extent is this effect a function of general language processing? It could be that participants who are quicker to activate L2 target words are also quicker in L1—They may be better at language processing in general, leading to greater L2 proficiency. Future work should examine whether and how speed of lexical activation correlates across languages, both for balanced bilinguals and for L2 learners.

However, the effect of proficiency was not just limited to early activation: we also saw correlations at later timepoints in lexical competition. There was a significant correlation of LexTALE-Esp score on the asymptote of targets (maximum looks; Figure 6B, C, & D) as well as cohorts (offset baseline; Figure 7B & C). These were reversed: more proficient listeners showed higher target asymptotes but lower cohort asymptotes. This is consistent with the idea that less proficient L2 listeners were struggling to resolve competition (and indeed, in the Spanish/Spanish condition as a group they were not able to do so). The strength of these correlations suggests that, at this stage of L2 learning, proficiency may be more closely related to the robustness of the lexicon and how well it is organized.

Importantly, differences in these later timepoints are not observed over the typical timecourse of L1 development; rather, this patterns similarly as individual differences in ability seen in kids with development language disorder (McMurray et al., 2014, McMurray et al., 2010). While we would not argue that L2 learners represent disordered language, it may be that the late onset of L2 acquisition leaves a similarly fragile lexicon to what children with DLD develop (for different reasons). It is possible that we see these differences in resolution more clearly among adults L2 learners in part because adults vary more widely in their ability to acquire a second language than typically developing monolingual children acquiring an L1. Alternatively, this poor resolution may reflect a less well-organized L2 lexicon in these adult learners due to individual differences in language ability.

We did not observe a correlation of early cohort activation (height of the cohort) and LexTALE-Esp score in either the Spanish-Spanish condition or the Spanish-English condition. Figure 7A shows that there may be some differences in fixations of the cohort between high performers and low performers, however, the difference was not significant. Visually, it appears to contrast our predictions based on the L1 literature (Figure 1A). However, Blumenfeld and Marian (2013) showed that more proficient bilinguals listening in their L1 show greater activation of the L2 cohort early as well as greater suppression of the L2 cohort later. Therefore, we might have reasonably expected that cohort activation in the within-L2 condition would reflect something similar—that more proficient L2 learners would show higher peak fixations to the cohort due to faster processing. This null result may derive from either—or both—of two causes.

First, this null result may be a function of level of language exposure. The L2 learners we tested were only in their second year of college Spanish. This is relatively “younger” in the developmental timeline of the L2, compared to the bilinguals in Blumenfeld and Marian (2013), who acquired their L2 around or before adolescence. However, cohort effects can be somewhat complicated, and they depend on the amount of phonological overlap between the competitors and as we described they may be driven largely by temporary ambiguity in the signal, not internal dynamics of processing.

Second, measuring cohort peak specifically is psychometrically challenging. In a test/retest reliability study of the VWP, cohort peak was relatively reliable, however, about a third of the variance in peak height was related to more general visual and attentional processes (unlike other properties which were more distinctly auditory-driven) (Farris-Trimble and McMurray, 2013). Therefore, it is possible that we were simply unable to detect a significant correlation of cohort peak height and proficiency in the present study and/or with the stimuli that we used.

Taken together, we see that proficiency correlates with eye-tracking parameters in several different ways, both at early and late timepoints in spoken word recognition. Our data suggests a mix of factors may influence lexical competition dynamics at this early stage of L2 learning—both in terms of variability in latent traits as well as differences in developmental progression and language experience. While the present study cannot parse these differences apart, the results do suggest a number of future directions. First, if latent trait type factors contribute differentially to spoken word recognition, then a larger battery of standardized tests would better capture and differentiate between the contributions of these factors in spoken word recognition. Correlating these scores with parameters of lexical competition dynamics could elucidate how these latent traits may be important, particularly for the later effects that we observe. Second, if language experience is the primary driver of the early correlations with proficiency, then it would be useful to study the effect of the immersive experience on spoken word recognition. For example, how would a summer or semester abroad affect the development within-L2 competition dynamics? And finally, cross-sectional or longitudinal work over the course of acquisition would be critical for charting the development through an L2 curriculum. This would give us a more comprehensive picture of the maturation of the L2 lexicon and could help inform classroom practices to bring about more robust learning.

Conclusions

Overall, our study shows that L2 learners build their new lexica in a way that enables immediate incremental processing—much like L1 listeners—and that they exhibit competition across both lexica. Critically, how these competition dynamics play out over time is related to proficiency in ways that implicates roles of both learning and development, as well as robustness of lexical organization and language ability. It also underscores the importance of developing realtime language processing skills for L2 learners and the need for examining bilinguals with a fuller range of abilities to create a more comprehensive picture of how these skills develop.

Acknowledgements

This work was supported by an Interdisciplinary Research Grant from the DeLTA (Developmental and Learning from Theory to Application) Center at the University of Iowa, awarded to MES. We would like to thank Jennifer Dibbern and Isaac Fisher for extensive help with stimulus selection and design. We would like to thank Hianca Pinho for help with hours of with data collection and assessment scoring.

Appendix A: Items used in the experiment

Table A: Item sets.							
Spanish-Spanish Cohort Pairs (unrelated words, when target is from Spanish-English)				Spanish-English Cohort Pairs (unrelated words, when target is from Spanish-Spanish)			
azúcar	sugar	***azul***	blue	***chanclas***	flipflops	*pizarra*	**chalkboard**
manzana	apple	***mantel***	tablecloth	***lado***	side	*armario*	**locker**
camiseta	t-shirt	***camas***	beds	***gato***	cat	*ajo*	**garlic**
cuchara	spoon	***cuchillo***	knife	***billetera***	wallet	*pelota*	**beachball**
gordo	fat	***gorra***	cap	***sangre***	blood	*fútbol*	**soccer**
señora	woman	***semana***	week	***fregadero***	sink	*congelador*	**freezer**
cielo	sky	***ciencia***	science	***botas***	boots	*frontera*	**border**
pato	duck	***papas***	potatoes	***caras***	faces	*alfombra*	**carpet**
abrazo	hug	***abrigo***	coat	***dormir***	sleep	*entrada*	**doorway**
cuello	neck	***cuento***	story	***silla***	chair	*mariscos*	**seafood**
estufa	stove	***estrella***	star	***plata***	silver	*jugador*	**player**
rojo	red	***ropa***	clothes	***concha***	shell	*esquina*	**corner**
espalda	back	***escuela***	school	***mar***	sea	*ratón*	**mouse**
pecho	chest	***perro***	dog	***leche***	milk	*clase*	**lecture**
aceite	oil	***asientos***	chairs	***pavo***	turkey	*fiesta*	**party**
vaso	cup	***vaca***	cow	***mundo***	world	*película*	**movie**
carpintero	woodworker	***cartas***	letters	***pastel***	cake	*bolsillo*	**pocket**
camarero	waiter	***camarones***	shrimp	***revistas***	magazine	*árbitro*	**referee**
estudiante	student	***estómago***	stomach	***caja***	box	*disfraz*	**costume**
desayuno	breakfast	***desierto***	desert	***falda***	skirt	*campesino*	**farmer**
almohadas	pillows	***almuerzo***	lunch	***tiburón***	shark	*dientes*	**teeth**
naranja	orange	***narcótico***	drugs	***pájaro***	bird	*cerámica*	**pottery**
caballos	horses	***cabello***	hair	***piña***	pineapple	*gente*	**people**
bocas	mouths	***boda***	wedding	***carne***	meat	*algodón*	**cotton**
hueso	bone	***huevos***	eggs	***pollo***	chicken	*puerta*	**port**
bebé	baby	***bebida***	drink	***fresa***	strawberry	*amigo*	**friend**
cerveza	beer	***cebolla***	onion	***pan***	bread	*olla*	**pot**
escritorio	desk	***escaleras***	stairs	***traje***	suit	*pista*	**track**
bolígrafo	pen	***boletos***	tickets	***mano***	hand	*limpiando*	**mopping**
dinero	money	***dios***	god	***lágrima***	tear	*medallón*	**locket**

References.

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**,** 419-439.

ANDERSON, M. C. & NEELY, J. H. 1996. Interference and inhibition in memory retrieval. *Memory.* Elsevier.

ANDREWS, S. & HERSCH, J. 2010. Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General,* 139**,** 299.

BHATIA, T. K. & RITCHIE, W. C. 2014. *The handbook of bilingualism and multilingualism*, John Wiley & Sons.

BLUMENFELD, H. K. & MARIAN, V. 2007. Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and cognitive processes,* 22**,** 633-660.

BLUMENFELD, H. K. & MARIAN, V. 2013. Parallel language activation and cognitive control during spoken word recognition in bilinguals. *Journal of Cognitive Psychology,* 25**,** 547-567.

BOERSMA, P. 2006. Praat: doing phonetics by computer. http://www.praat.org/.

BUILDER, S. R. E. 2011. SR Research Experiment Builder 1.10.165. Missassauga, Ontario, Canada: SR Research Ltd,.

CHATEAU, D. & JARED, D. 2000. Exposure to print and word recognition processes. *Memory & cognition,* 28**,** 143-153.

CHEN, H.-C. & HO, C. 1986. Development of Stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 12**,** 397.

CLARK, H. H. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior,* 12**,** 335-359.

CUTLER, A., WEBER, A. & OTAKE, T. 2006. Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics,* 34**,** 269-284.

DAHAN, D., MAGNUSON, J. S. & TANENHAUS, M. K. 2001a. Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive psychology,* 42**,** 317-367.

DAHAN, D., MAGNUSON, J. S., TANENHAUS, M. K. & HOGAN, E. M. 2001b. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes,* 16**,** 507-534.

DAVIS, M. H. & GASKELL, M. G. 2009. A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 364**,** 3773-3800.

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**,** 843-863.

DOLLAGHAN, C. 1998. Spoken word recognition in children with and without specific language impairment. *Applied Psycholinguistics,* 19**,** 193-207.

FARRIS-TRIMBLE, A. & MCMURRAY, B. 2013. Test–retest reliability of eye tracking in the Visual World Paradigm for the study of real-time spoken word recognition. *Journal of Speech, Language, and Hearing Research*.

FERNANDES, T., KOLINSKY, R. & VENTURA, P. 2009. The metamorphosis of the statistical segmentation output: Lexicalization during artificial language learning. *Cognition,* 112**,** 349-366.

FRAUENFELDER, U. H., SCHOLTEN, M. & CONTENT, A. 2001. Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. *Language and cognitive processes,* 16**,** 583-607.

HANNAGAN, T., MAGNUSON, J. S. & GRAINGER, J. 2013. Spoken word recognition without a TRACE. *Frontiers in psychology,* 4**,** 563.

HENDRICKSON, K., GOODWIN, C., KLEIN, K. & MCMURRAY, B. submitted. The profile of real-time competition in spoken and written word recognition: more similar than different. *Quarterly Journal of Experimental Psychology*.
IZURA, C., CUETOS, F. & BRYSBAERT, M. 2014. Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica,* 35**,** 49-66.
JONIDES, J. & NEE, D. E. 2006. Brain mechanisms of proactive interference in working memory. *Neuroscience,* 139**,** 181-193.
JU, M. & LUCE, P. A. 2004. Falling on Sensitive Ears: Constraints on Bilingual Lexical Activation. *Psychological Science,* 15**,** 314-318.
KAPNOULA, E. C. & MCMURRAY, B. 2016. Newly learned word forms are abstract and integrated immediately after acquisition. *Psychonomic bulletin & review,* 23**,** 491-499.
KAPNOULA, E. C., PACKARD, S., GUPTA, P. & MCMURRAY, B. 2015. Immediate lexical integration of novel word forms. *Cognition,* 134**,** 85-99.
LEACH, L. & SAMUEL, A. G. 2007. Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology,* 55**,** 306-353.
LEE, I. & PREACHER, K. 2016. Calculation for the test of the difference between two dependent correlations with one variable in common Available from http://quantpsy. org. *Accessed Sept*.
LEMHÖFER, K. & BROERSMA, M. 2012. Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods,* 44**,** 325-343.
LINDSAY, S. & GASKELL, M. G. 2013. Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 39**,** 608.
LUCE, P. A. & PISONI, D. B. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and hearing,* 19**,** 1.
MAGILL, R. A. & HALL, K. G. 1990. A review of the contextual interference effect in motor skill acquisition. *Human movement science,* 9**,** 241-289.
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**,** 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**,** 202.
MARIAN, V. & SPIVEY, M. 2003. Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition,* 6**,** 97-115.
MARSLEN-WILSON, W. & ZWITSERLOOD, P. 1989. Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human perception and performance,* 15**,** 576.
MARSLEN-WILSON, W. D. 1987. Functional parallelism in spoken word-recognition. *Cognition,* 25**,** 71-102.
MCCLELLAND, J. L. & ELMAN, J. L. 1986. The TRACE model of speech perception. *Cognitive psychology,* 18**,** 1-86.
MCCLOSKEY, M. & COHEN, N. J. 1989. Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of learning and motivation.* Elsevier.
MCMURRAY, B. 2017. EyelinkAnal.
MCMURRAY, B. 2020. Nonlinear Curvefitting for Psycholinguistic (and other) Data. *osf.io/4atgvg*.
MCMURRAY, B., DANELZ, A., RIGLER, H. & SEEDORFF, M. 2018. Speech categorization develops slowly through adolescence. *Developmental psychology,* 54**,** 1472.
MCMURRAY, B., ELLIS, T. P. & APFELBAUM, K. S. 2019a. How Do You Deal With Uncertainty? Cochlear Implant Users Differ in the Dynamics of Lexical Processing of Noncanonical Inputs. *Ear and hearing,* 40**,** 961-980.
MCMURRAY, B., KAPNOULA, E. C. & GASKELL, M. G. 2016. Learning and integration of new word-forms: Consolidation, pruning, and the emergence of automaticity. *Speech perception and spoken word recognition.* Psychology Press.

MCMURRAY, B., KLEIN-PACKARD, J. & TOMBLIN, J. B. 2019b. A real-time mechanism underlying lexical deficits in developmental language disorder: Between-word inhibition. *Cognition,* 191, 104000.
MCMURRAY, B., MUNSON, C. & TOMBLIN, J. B. 2014. Individual differences in language ability are related to variation in word recognition, not speech perception: Evidence from eye movements. *Journal of Speech, Language, and Hearing Research,* 57, 1344-1362.
MCMURRAY, B., SAMELSON, V. M., LEE, S. H. & TOMBLIN, J. B. 2010. Individual differences in online spoken word recognition: Implications for SLI. *Cognitive psychology,* 60, 1-39.
MEARA, P. & JONES, G. 1987. Tests of vocabulary size in English as a foreign language. *Polyglot,* 8, 1-40.
MEARA, P. & JONES, G. 1990. 1990: The Eurocentres Vocabulary Size Test. 10K. Zurich: Eurocentres.
MUÑOZ, C. & SINGLETON, D. 2011. A critical review of age-related research on L2 ultimate attainment. *Language teaching,* 44, 1.
MUYSKEN, P. 2013. Two linguistic systems in contact: Grammar, phonology, and lexicon. *The handbook of bilingualism and multilingualism*, 193-216.
PRESTON, M. S. & LAMBERT, W. E. 1969. Interlingual interference in a bilingual version of the Stroop color-word task. *Journal of Verbal learning and Verbal behavior,* 8, 295-301.
QU, Q., CUI, Z. & DAMIAN, M. F. 2018. Orthographic effects in second-language spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 44, 1325.
RIGLER, H., FARRIS-TRIMBLE, A., GREINER, L., WALKER, J., TOMBLIN, J. B. & MCMURRAY, B. 2015. The slow developmental time course of real-time spoken word recognition. *Developmental psychology,* 51, 1690.
ROEMBKE, T. C., HAZELTINE, E., REED, D. K. & MCMURRAY, B. 2019. Automaticity of word recognition is a unique predictor of reading fluency in middle-school students. *Journal of Educational Psychology,* 111, 314.
ROHRER, D. 2012. Interleaving helps students distinguish among similar concepts. *Educational Psychology Review,* 24, 355-367.
SEEDORFF, M., OLESON, J. & MCMURRAY, B. 2018. Detecting when timeseries differ: Using the Bootstrapped Differences of Timeseries (BDOTS) to analyze Visual World Paradigm data (and more). *Journal of Memory and Language,* 102, 55-67.
SEGALOWITZ, N. 2010. *Cognitive bases of second language fluency*, Routledge.
SEKERINA, I. A. & BROOKS, P. J. 2007. Eye movements during spoken word recognition in Russian children. *Journal of Experimental Child Psychology,* 98, 20-45.
SHEA, J. B. & MORGAN, R. L. 1979. Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental psychology: Human Learning and memory,* 5, 179.
SPIVEY, M. & MIRMAN, D. 2001. Retroactive interference in neural networks and in humans: the effect of pattern-based learning. *Connection Science,* 13, 257-275.
SPIVEY, M. J., GROSJEAN, M. & KNOBLICH, G. 2005. Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences,* 102, 10393-10398.
SPIVEY, M. J. & MARIAN, V. 1999. Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological science,* 10, 281-284.
TANENHAUS, M. K., SPIVEY-KNOWLTON, M. J., EBERHARD, K. M. & SEDIVY, J. C. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science,* 268, 1632-1634.
TZELGOV, J., HENIK, A. & LEISER, D. 1990. Controlling Stroop interference: Evidence from a bilingual task. *Journal of experimental psychology: Learning, memory, and cognition,* 16, 760.

VEIVO, O., JÄRVIKIVI, J., PORRETTA, V. & HYÖNÄ, J. 2016. Orthographic activation in L2 spoken word recognition depends on proficiency: Evidence from eye-tracking. *Frontiers in Psychology,* 7, 1120.

VEIVO, O., PORRETTA, V., HYÖNÄ, J. & JÄRVIKIVI, J. 2018. Spoken second language words activate native language orthographic information in late second language learners. *Applied Psycholinguistics,* 39, 1011-1032.

WEBER, A. & CUTLER, A. 2004. Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language,* 50, 1-25.

YAP, M. J., BALOTA, D. A., TSE, C.-S. & BESNER, D. 2008. On the additive effects of stimulus quality and word frequency in lexical decision: evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 34, 495.

ZHANG, X. & SAMUEL, A. G. 2018. Is speech recognition automatic? Lexical competition, but not initial lexical access, requires cognitive resources. *Journal of Memory and Language,* 100, 32-50.

ZHAO, L., YUAN, S., GUO, Y., WANG, S., CHEN, C. & ZHANG, S. 2020. Inhibitory control is associated with the activation of output-driven competitors in a spoken word recognition task. *The Journal of General Psychology*, 1-28.