Language, Cognition and Neuroscience

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

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

McCall E. Sarrett, Christine Shea & Bob McMurray

LANGUAGE, COGNITION AND NEUROSCIENCE
2022, VOL. 37, NO. 2, 165–181
https://doi.org/10.1080/23273798.2021.1952283

REGULAR ARTICLE

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

McCall E. Sarrett[a], Christine Shea[b] and Bob McMurray[c]

[a]Dept. of Psychological & Brain Sciences, Villanova University, Villanova, PA, USA; [b]Dept. of Spanish and Portuguese, Dept. of Linguistics, University of Iowa, Iowa City, IA, USA; [c]Dept. of Psychological and Brain Sciences, Dept. of Communication Sciences and Disorders, Dept. of Linguistics, University of Iowa, Iowa City, IA, USA

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 a 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).

ARTICLE HISTORY
Received 22 October 2020
Accepted 29 June 2021

KEYWORDS
Second language acquisition; spoken word recognition; lexical competition; language proficiency; visual world paradigm

Introduction

In an increasingly globalised world, people often acquire second and third languages. This is not trivial. A great deal of both pedagogical and scientific work focuses on how learners acquire language *knowledge* (i.e. phonology, words, and grammar; see Muysken, 2013 for a review). Less recognised is that second language (L2) learners must also be able to rapidly access this new knowledge in real-time, given that the ultimate goal of L2 acquisition is the ability to communicate and comprehend ideas and information efficiently. This is particularly true for adult L2 learners who are acquiring L2 knowledge post-adolescence and primarily in classroom contexts (who 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 this efficiency develops in adult L2 learners.

Lexical competition. Spoken word recognition is fundamental for L2 learners, because words link sound to syntactic and semantic knowledge, and ultimately to conceptual structures that are not language specific. Recognising 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 recognise words.

One fundamental challenge for 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,* when only *chee-* (/tʃi/) has been heard, 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 (Allopenna et al., 1998 Marslen-Wilson & Zwitserlood, 1989;). These candidates compete for recognition (Dahan et al., 2001b; Luce & Pisoni, 1998; Marslen-Wilson, 1987). As more acoustic information arrives, one candidate wins out as the recognised word. This process of lexical competition is critical for efficient and accurate perception of spoken words (Hannagan et al., 2013; Marslen-Wilson, 1987 Mcclelland & Elman, 1986;). For bilingual listeners, lexical competition may be vastly more complicated, as competition unfolds over multiple lexica.

The most prominent recent methodology for examining lexical competition is the Visual World Paradigm (VWP), which can track lexical competition in real time (Allopenna et al., 1998; Tanenhaus et al., 1995). In a typical VWP task, participants see a set of pictures which represent potential lexical candidates that may

CONTACT McCall E. Sarrett mccall.sarrett@villanova.edu Dept. of Psychological & Brain Sciences, Villanova University, Tolentine Hall 334, 800 E Lancaster Ave, Villanova, PA 19085, USA
Supplemental data for this article can be accessed https://doi.org/10.1080/23273798.2021.1952283.

be considered (and potentially ruled out) when a 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 candidates as speech unfolds in time.

For example, one might present a target word, *cheeseburger,* along with its own picture, a picture of a *cheerleader* (a cohort, which overlaps with the target initially), and a *grasshopper* (an unrelated word). As the participant hears *cheeseburger*, both *cheeseburger* and *cheerleader* are fixated equally at first, but as more acoustic information arrives, *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 monolingual listeners. The VWP has shown how real time spoken word recognition in the L1 develops in school age children (Rigler et al., 2015), and how it is related to language outcomes (McMurray et al., 2010; McMurray et al., 2014). However, less is known about its development in adult L2 learners. Understanding competition dynamics in this group may be important for two reasons. First, as we describe, broader work in bilingualism (reviewed below) has used lexical competition to ask questions about the degree to which the lexica in each language are modular or encapsulated. This has not yet been addressed in 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 listener knows the language 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 an entire lexicon, the two lexica could interact, or there may be a single lexicon comprising words from both languages. In these cases, when hearing *chief,* listeners should consider Spanish competitors like *chicle.* This raises an additional cognitive challenge as multilinguals must manage competition among a much larger pool of words (than monolinguals). Nevertheless, such flexibility could be useful for multilinguals who regularly code switch between languages, or must regularly communicate in situations where the language of a speaker or of a conversation cannot be is not known in advance. In these cases, activating lexical competitors across multiple languages could lead to more flexible spoken word recognition.

Several studies have addressed this debate in multilinguals. Much of this work has utilised a cross-linguistic VWP (see also Chen & Ho, 1986; Preston & Lambert, 1969; Tzelgov et al., 1990 for evidence with a Stroop task). 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 increased fixations to an English cohort (*fish*) (Spivey & Marian, 1999). This pattern of cross-language activation has been since shown in bilinguals of many languages: Dutch-English (Weber & Cutler, 2004), Spanish-English (Ju & Luce, 2004), Japanese-English (Cutler et al., 2006), Finnish-French (Veivo et al., 2018), and German-English (Blumenfeld & Marian, 2007). Thus, cross-language competition seems to be a general feature of word recognition in balanced bilinguals.

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 & Marian, 2007; Cutler et al., 2006 Ju & Luce, 2004; Marian & Spivey, 2003;; Veivo et al., 2018; Weber & Cutler, 2004) have relied largely on participants who learned their L2 early (in childhood or during adolescence) and have spent significant time in immersive language settings. As a result, they 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 acquired concurrently. Drawing from the learning and memory literature, this mimics "interleaved" exposure (i.e. a "back-and-forth" pattern of exposure between alternating languages), which may result in a greater degree of interactivity

(Rohrer, 2012). In contrast, adult L2 learners already have an established L1 and are acquiring an L2 later (post-adolescence) and via primarily classroom-based contexts. The latter mimics a "blocked" exposure (i.e. one language is established, and the next is acquired, sequentially), which tends to lead to less robust learning and is more susceptible to both proactive and retroactive interference (Anderson & Neely, 1996 Jonides & Nee, 2006; Mccloskey & Cohen, 1989; Spivey & Mirman, 2001). It may also lead to a differently organised lexicon. As such, it may not be safe to assume that L2 learners will show the same cross-language competition as balanced bilinguals.

Second, if adult L2 learners do show cross-linguistic lexical competition, their ability to manage this competition may be important. L2 learners must be able to efficiently suppress the much more established L1 competitors to allow L2 word recognition to proceed. This suggests a key role for L2 proficiency, especially in their early stages of acquisition. As we describe, work on spoken word recognition in L1 acquisition offers a useful framing that may help give insight to the role of proficiency on real-time processes.

Lexical competition and proficiency

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

First, while lexical competition is strongly driven by the unfolding acoustic input (and the temporary ambiguity that it creates), the dynamics of competition are also related to other internal factors such as the *structure of the lexicon*. These dynamics are likely to change as people acquire new words in their L2. 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 recognised more slowly than words with lower cohort densities (Magnuson et al., 2007). Moreover, 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 word occurs. And of course, frequency is a proxy for how well-learned a word is. Such factors could broadly be construed in the context of lexical quality (Perfetti & Hart, 2002), the robustness of the lexical representation and the strength of the connections between phonemes and corresponding lexical entries. Such factors may influence the speed with which a word is recognised, the strength of competitor consideration, and the degree to which they are suppressed. Thus, the timecourse of lexical competition is influenced by the number of competitors in the lexicon, how well learned they are, and the general pattern of connectivity. All of these are likely to be dramatically altered by acquiring a second language.

Second, work on *word learning* in monolinguals 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., 2001a). Recent work has shown that such inhibitory connections are created very rapidly when learning new words, and are present even on the initial day of learning (Kapnoula et al., 2015; Kapnoula & McMurray, 2016; and see Fernandes et al., 2009; Lindsay & Gaskell, 2013) thus these same processes are likely to be available in the L2 lexicon.

Finally, 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 (see 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-olds are slower to fully activate target words than 16-year-olds, and they show higher competition at onset (Rigler et al., 2015; see also McMurray et al., 2018; Sekerina & Brooks, 2007). That is, typical development helps listeners extract more information from the signal at each time point. Second, within an age (in this case, adolescents), 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., 2010; McMurray et al., 2014; McMurray et al., 2019b; see also Roembke et al., 2019 for work with proficiency and written word recognition). These studies all suggest that the real-time dynamics of spoken word recognition are broadly shaped by proficiency. This is reflected both in development as listeners become more skilled and have increasing language experience, and by individual differences in language ability within an age group. Both may be construed as different forms or consequences of lexical robustness,

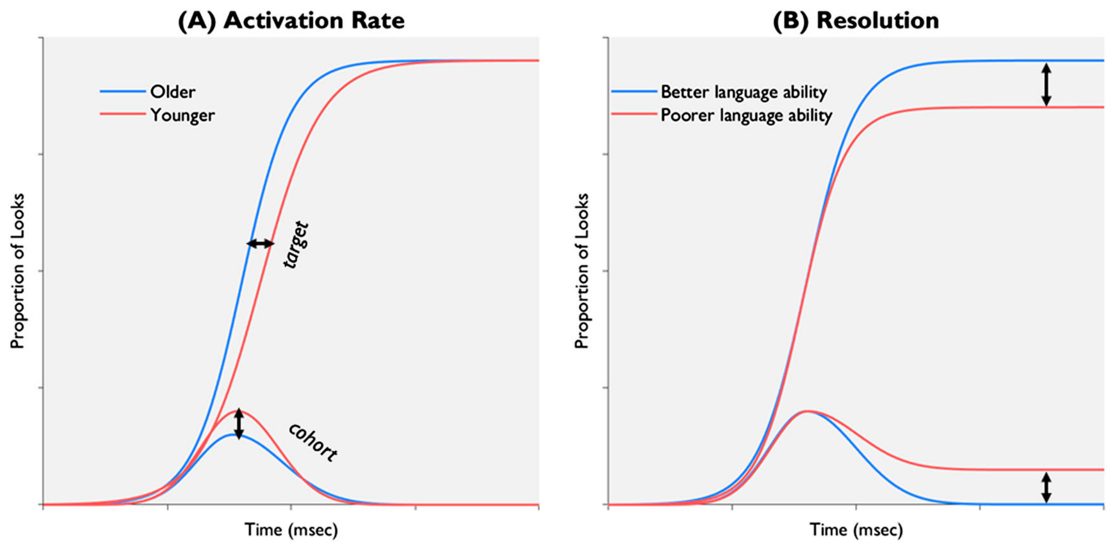

Figure 1. A schematic showing data patterns from L1 literature that may inform predictions for L2 acquisition. (A) Differences in activation rate result from the timecourse of development between younger and older children. If developing L2 proficiency is more dependent on amount of language exposure, we may expect L2 proficiency to correlate more highly with these early measures of lexical competition: timing of looks to the target and cohort height. (B) Differences in resolution due to individual language ability and robustness of the lexicon. If developing L2 proficiency is more dependent on individual language ability, we may expect L2 proficiency to correlate more instead with these later measures of competition resolution: maximum asymptote for targets and offset baseline for cohorts.

suggesting this construct translate in multiple ways to real-time processing.

For multilinguals, there is less evidence on the nature of the relationship between L2 proficiency and real-time processing (than in monolinguals). 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 (similar to written word recognition in bilinguals: Veivo et al., 2016; Veivo et al., 2018; see also Qu et al., 2018). Even balanced bilinguals have not yet been fully characterised within either of their languages along fundamental dimensions like activation rate or resolution. The existing literature with balanced bilinguals does suggest that language proficiency is associated with general aspects of processing efficiency, but as of yet, we do not know how these patterns may or may not relate to adult L2 acquisition.

There has been no work on adult 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 presumably more fragile, and may be particularly true in the case of L2 classroom learners, who acquire the target language in a limited context where it is not used on a daily basis.

The L1 work raises two clear loci of L2 proficiency effects in the timecourse of lexical competition: activation rate and resolution. 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; McMurray et al., 2018; Rigler et al., 2015; Sekerina & Brooks, 2007). In this case, 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 organisation, 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., 2019b, 2010, 2014). 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 unique progression (relative to the L1) could result in either slower development (predicting differences in activation rates) or less robust language representations (predicting differences in resolution), or both. 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 present study

The present study had two goals. First, we asked if adult L2 learners show both within-language competition in their L2 and cross-language competition between their L2 and L1. Second, we asked how individual differences in language proficiency correlate with two components of real-time word recognition: early activation and later resolution of competitors (both within-L2 and cross-language).

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 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 similar between-L2/L1 competition in adult L2 learners as has been observed in balanced bilinguals. If, however, the developmental progression in adult L2 learners leads to lexica that are differently organised (i.e. modular lexica, due to a "blocked" language exposure), then we may not see between-L2/L1 competition.

With respect to the second question concerning proficiency, 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 estimated different properties of the timecourse of word recognition by fitting curves to the fixations for each participant. This yielded a set of parameters that capture, for example, how quickly participants fixate the target or how fully they suppress competitors, that can then be correlated with 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 in Spanish (e.g. study abroad) were ineligible. 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, 27 female; approximate age range: 18–24 years old).

Design and items

We used a cross-linguistic version of the Visual World Paradigm based on Marian and Spivey (2003). Participants saw a set of four pictures and heard a Spanish word. Participants chose which of the 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 an 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.

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 the first phonemes of the Spanish word (e.g. *botas* [ˈbotas] "boots") overlapped with an n English word ("border" [ˈbɔrdər]), and not with its Spanish translation (*frontera*). This condition averaged 2.1 phonemes (range: 1-3, average: 2.1) of overlap. 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.

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 are generally stressed on the first syllable in both languages. If it was not possible to find word pairs of the same syllable length (due to the need for familiarity by early learners), then stress pattern was controlled within a pair such that the location of primary stress was roughly consistent. Full item sets as well as a further characterisation of stimulus parameters are given in Supplement S1.

This design resulted in three conditions (see Figure 2). First, 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. Second, 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. Finally, the No Competitor condition also came from the Spanish-English pair, where *frontera* ("border") was the target word. In this condition, *frontera* does not 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 (see Supplement S1). Each item within a set was repeated 4 times as the target word. This yielded 480 trials (30 sets × 4 items per set × 4 repetitions). Each item set consisted of one Spanish-Spanish cohort pair and one Spanish-English cohort pair. Both items in a Spanish-Spanish pair had a "reciprocal" competitor relationship (that is, we could test activation for *cielo* given *ciencia,* and for *ciencia* given *cielo*). Consequently, there were 240 trials in the Spanish-Spanish condition. In contrast, only one item of the Spanish-English pair had the specified competitor relationship (we could test activation for *frontera* "border", given *botas,* but when hearing *frontera*, there was no competitor). Thus, there were only 120 trials for each the Spanish-English competition as well as the No Competitor condition. Items occurred in each of the four corners of the screen on an equal numbers of trials.

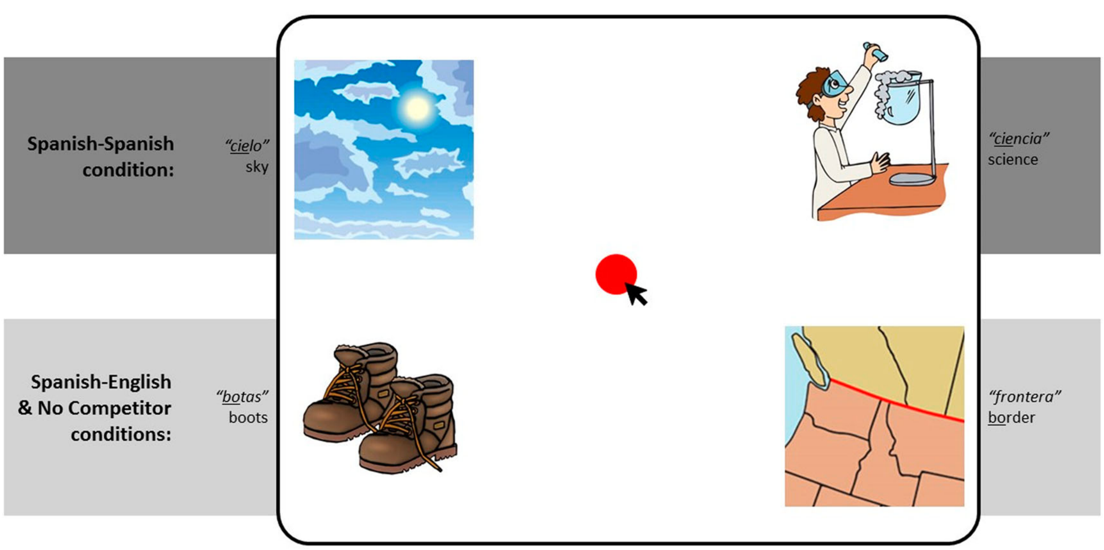

Figure 2. Example screen from the cross-linguistic 4AFC Visual World Paradigm task. Item sets consisted of one Spanish-Spanish competitor pair (top; e.g. cielo-ciencia) and one Spanish-English competitor pair (bottom; e.g. botas-border). This resulted in three conditions: Spanish-Spanish condition (cielo is the target, 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).

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 viewed by focus groups of native English speakers who selected the best image and identified any necessary modifications. Pictures were then edited to ensure canonical orientations and colours 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 normalised 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.

Procedure

Participants were greeted in English and all instructions were given in English. Participants gave informed consent according to IRB protocols at the University of Iowa and completed a short language background and demographic questionnaire. Participants then completed the LexTALE-Esp, followed by the Spanish VWP. The full experimental session took approximately one hour.

LexTALE-Esp. The original Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer & Broersma, 2012) was developed from the Eurocentres Vocabulary Size Test (EVST; Meara & Jones, 1987, 1990). It has since been adapted for 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 mark the items they believe to be words in that language.

The LexTALE-Esp is scored using a signal detection theory approach. This takes into account the number of words correctly identified, the nonwords correctly rejected, 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 neighbourhood priming task; Andrews & Hersch, 2010, written word recognition; Chateau & Jared, 2000, word identification times; Diependaele et al., 2013, lexical decision task performance; Yap et al., 2008)

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").

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, self-paced familiarisation, in which they saw items' pictures and names (orthographically), in Spanish only.

On each trial, four images appeared in the corners of a 1280 × 1024 pixel CRT monitor, 50 pixels away from the edges of the monitor. A small red dot (60-pixel diameter) appeared in the centre of the screen, which turned blue after 1000 msec, signalling the start of the trial. During this forced pre-scan period, participants could inspect the pictures to identify the visuo/semantic features 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 inter-trial interval of 250 msec before the screen advanced.

Eye-tracking methods

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. These early looks occurred in approximately 6.5% of trials, and were re-coded as looks to nothing (i.e. outside of any area of interest on the screen).[1] 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.

Quantifying lexical competition

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, the timing of the 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 minimises 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 goodness of the fit by computing the correlation between the data and the estimated curve, and by visually comparing the estimated curve to the 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 gradient descent method stopping at a local minimum; in this case, refitting from an alternative starting point was usually sufficient to ensure that the curve reflected the best possible fit. Only a few curves had to be refit for each condition 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).

One concern was that number of trials differed across conditions; this could lead to more robust data in the Spanish-Spanish than the Spanish-English conditions. Though it was not our intention to directly to compare them, the fact that the goodness of fits were similarly high for all conditions and all item types, suggest that summary statistics derived from these fits were equally valid in both conditions.

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 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; $F(2, 96) = 1.439$, $p = .242$).

Within-L2 lexical competition in L2 learners

Figure 3(A) shows the proportion of fixations to each of the three candidates over time in the Spanish-Spanish

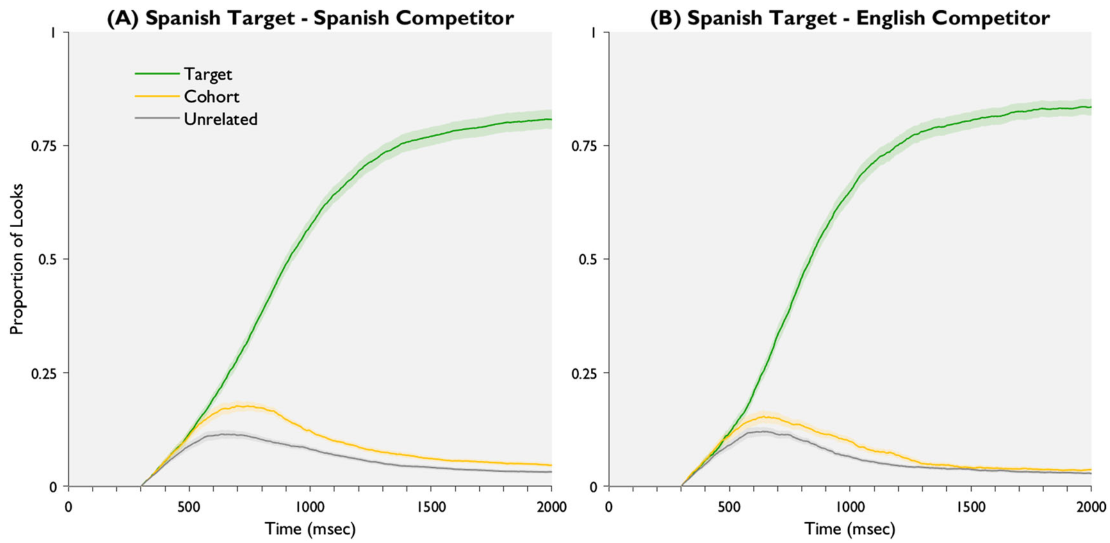

Figure 3. (A) Proportion of looks over time as a function of item type in the within-Spanish condition. Participants showed significantly more peak looks to the Spanish cohort compared to the unrelated. Participants also showed a significantly higher offset baseline in the cohort condition compared to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

condition. The unrelated curve is calculated as an average of looks to either of the two unrelated items 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. Figure 3(B) 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.

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? 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 ($t_1(32) = 10.42$, $p < .0001$; $t_2(29) = 6.83$, $p < .0001$). This pattern is remarkably consistent with what one would expect for monolinguals. 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? If so, this would suggest that L2 learners show a similar consideration of words in both lexica; if not, it may indicate that L2 learners have strategies for functionally encapsulating or shutting off their L1 lexicon.

Figure 3(B) shows clear evidence for cross-language competition. 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 ($t_2(29) = .95$, $p = .35$). Thus, 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).

We also compared the magnitude of within- and between-language competition (see Supplement S2). While these analyses should be interpreted with caution as we only used a small subset of well matched items, they suggest the magnitude of within- and between-language competition is similar, although English competitors are active slightly earlier and suppressed somewhat more fully.

An alternative account

One alternative explanation for the cross-linguistic effect 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, subjects may have implicitly noticed the association within the Spanish-Spanish pair and reasoned by extension that the other two Spanish-English items may be as well. This could lead to enhanced English cohort looking, even without any true competition.

To address this, we examined 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 (e.g. participants only heard *frontera*, which does not share any phonological overlap with *botas*). If the effect were driven purely by associations between pictures during the experiment, we would expect to see a significant increase in looks to *botas* when hearing *frontera*.

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: $t_1(32) = 0.17$, $p = .864$; $t_2(29) = 0.42$, $p = .67$).

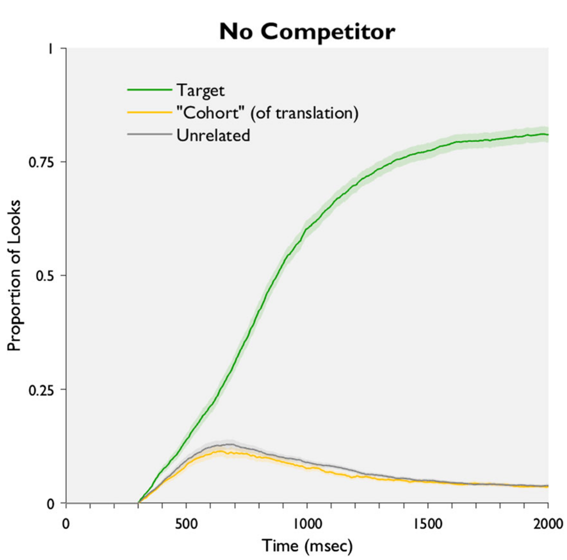

Figure 4. Proportion of looks over time in the No Competitor condition. Peak looks to the cohort of the translation were not significantly different when compared to the unrelated. Offset baseline also did not differ between the two conditions.

This indicates that the cross-language cohort effect is not due to subjects having learned associations between the pictures. Moreover, this also suggests that L2 learners are not explicitly translating items on the screen into their L1. If they were, we would expect to see spreading activation of the "cohort" of the translation in this condition, over and above that of the baseline.

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. We correlated the continuous LexTALE-Esp scores across participants with several key parameters from the eyetracking data to capture the full range of performance. Our first analyses focused on the target fixations. We started with a simple data visualisation, using a median split of LexTALE-Esp scores to examine overall pattern of differences. The median split was used for data visualisation purposes only. L2 learners with a LexTALE-Esp score of 7 or below were included as the low performance group ($N = 17$; shown in red in Figures 5 and 6); those above a 7 were classified in the high performance group ($N = 16$; shown in blue in Figures 5 and 6). This data visualisation (shown in Figure 5A) suggested robust differences in the target fixations between lower and higher proficiency L2 learners. We quantified this statistically by correlating continuous LexTALE proficiency to the activation rate of the target word.

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 log-scaled the slope and then computed the Z-score for crossover and slope separately. The crossover Z-score was then multiplied by −1 (since larger slopes mean a faster-rising curve, but later crossovers mean a slower-rising curve), and 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.

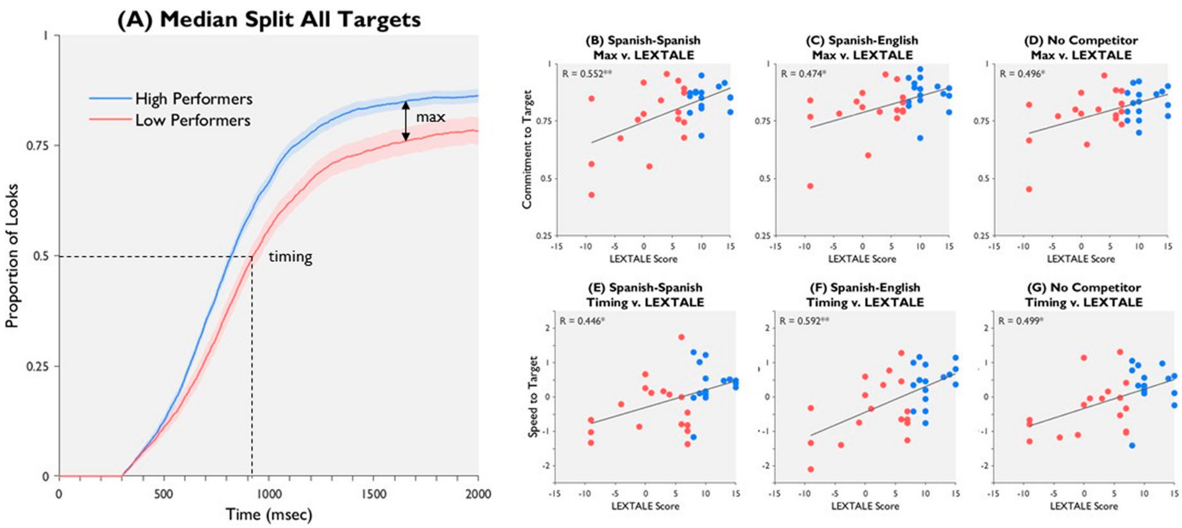

Figure 5. (A) Average looks to the target in all conditions, split by high (blue) and low (red) performers on LexTALE-Esp. The top row of scatterplots shows maximum looks to target in each condition vs. LexTALE-Esp: (B) Spanish-Spanish; (C) Spanish-English; (D) No Competitor. The bottom row of scatter plots shows timing in each condition vs LexTALE-Esp: (E) Spanish-Spanish; (F) Spanish-English; (G) No Competitor.

Figure 5(A), shows average looks to target across all conditions, visualised by high performers (blue) and low performers (red). The top row of scatter plots (Figure 5B, C, and D) show the correlation of the continuous LexTALE-Esp score with maximum target fixations across different conditions. We found that LexTALE-Esp significantly correlated with the maximum in all three conditions: (5B) Spanish-Spanish $R = 0.552$,

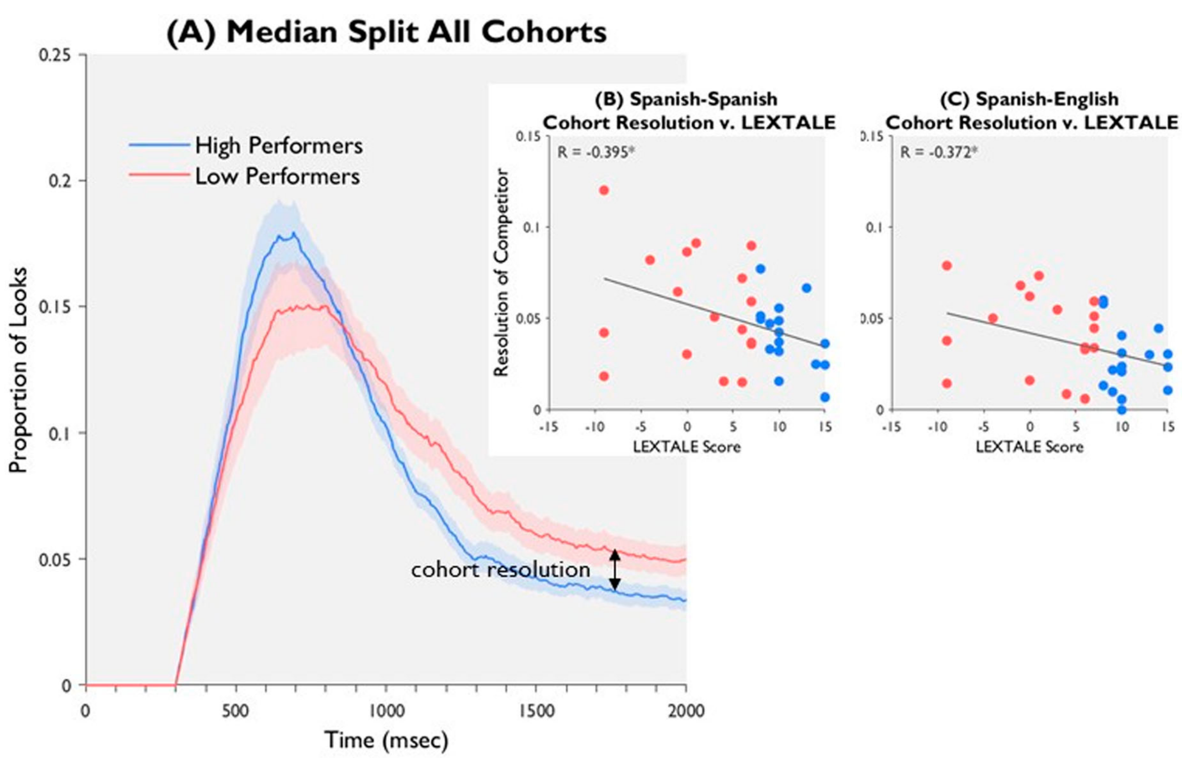

Figure 6. (A) Average looks to the cohort across conditions, split by high (blue) and low (red) performers on LexTALE-Esp. The subplots show correlation of cohort resolution vs. LexTALE-Esp score in (B) the Spanish-Spanish condition and (C) the Spanish-English condition.

t(31) = 3.689, $p < .001$; (5C) Spanish-English $R = 0.474$, t(31) = 3.001, $p = .005$; (5D) No Competitor $R = 0.496$, t(31) = 3.177, $p = .003$. L2 learners who scored better on the test of vocabulary and proficiency have higher maximum asymptotes.

The bottom row of scatter plots (Figure 5E, 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 & Preacher, 2016). Overall, L2 learners who had better vocabulary proficiency, as measured by LexTALE-Esp, also were significantly quicker to look to the target word.

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 largely 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 6(A) shows average looks to the cohort across all conditions, visualised by high performers (blue) and low performers (red). The two insets (Figure 6B and C) show significant correlations of LexTALE-Esp score with cohort resolution in both conditions, measured by the offset baseline (Figure 6B, left insert: $R = -0.395$, $t(31) = 2.396$, $p = .023$) and the English competitor condition (Figure 6C, right insert: $R = -0.372$, $t(31) = 2.232$, $p = .033$). These correlations were not significantly different between conditions ($z = -0.187$, $p = .85$). L2 learners with higher proficiency 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 an obligatory processing of the unfolding auditory input.

Discussion

This study sought to characterise 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 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.

Within-L2 competition

Even adult L2 learners who are relatively early in their Spanish acquisition show typical incremental processing in their L2 (Figure 3A). This is despite limited experience with the language (and for our participants, no immersive experience). This very rapid development of lexical competition fits with the current literature on word learning in L1, which suggests that new words are immediately integrated into an interconnected lexicon even on the same day as they are learned (Kapnoula et al., 2015; Kapnoula & McMurray, 2016; Magnuson et al., 2003). Thus, this may indicate that even when building a full lexicon of a new language, a similar rapid establishment of an interconnected lexical network is possible.

In general, the overall pattern of within-L2 lexical competition dynamics showed a similar pattern to that of monolingual word recognition. However, there was one notable difference. The offset asymptote of the L2 cohort never returned to baseline (it remained significantly greater than the unrelated). This suggests that lexical competition may not resolve as fully in adult L2 learners (though the degree of this resolution was related to the learners' proficiency). Functionally, this means that adult L2 learners may exhibit greater uncertainty with regards to which lexical item should have been activated, even at these late points in 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). A number of such mechanisms are thought to aid more complete competition resolution: lateral inhibition among words (Dahan et al., 2001a), decay (McMurray et al., 2010), mismatch between the input and the competitor (Frauenfelder et al., 2001), and possibly domain-general cognitive control (Zhang & Samuel, 2018; Zhao et al., 2020). It is possible that one or some of these factors, such as decay or lateral inhibition, could be different in the less robust lexical networks of adult L2 learners, and thus could contribute to this incomplete resolution.

Cross-linguistic competition

We also found clear evidence for cross-linguistic competition, between L2 and L1 (Figure 3B). L2 learners do not seem to have an explicit strategy for suppressing the irrelevant lexicon (i.e. their L1, given that the entire VWP task was in their L2). This suggests that despite their limited exposure to their L2, our participants' L2 and L1 lexica were still functionally interleaved.

We see a few possible mechanistic accounts for this. One possibility is that word recognition processing is not actually separated by language, that is, there may not exist multiple lexica in bilinguals. Instead, there could 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. Another possibility is that there are truly separate lexica, but that the phenomenon of cross-linguistic competition is driven primarily by the 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 acoustic signal at all times, this effect would still be observed, regardless of whether the lexica are 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 either 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). This is supported by analyses in the Supplement S2 which suggest that (at least in a limited set of items) when the amount of phonemic overlap is controlled, both within- and cross-language competition shows a fairly similar profile (although English competitors may be active earlier, and later be suppressed more fully). 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). Alternatively, a more pointed way to parse out these cross-linguistic links would be to utilise a cross-linguistic subphonemic mismatch paradigm (Dahan et al., 2001a). 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 connections via lexical inhibition.

Individual differences in proficiency

Finally, we asked whether individual differences in proficiency are related to the dynamics of within- and across-language competition (Figures 5 and 6). We found strong correlations between LexTALE-Esp score and many of the key parameters of lexical competition. There were significant correlations both early in the timecourse in the rate at which candidates became active (patterning with L1 development), and later in the timecourse in the degree to which targets were fully active and competitors fully suppressed (patterning with robustness of the lexicon). We discuss implications of the key findings below.

First, more proficient L2 learners recognised target words more quickly, in all three conditions (Figure 5E, F, and G). This tight link between activation rate and proficiency 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 may 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. 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 a participant does not know the target word on a given trial, they could instead deduce it from which items the word is *not*. This would delay the timing of looks to the target (even if they eventually respond correctly). However, if this were the case, we would expect looks to the target word to be more severely delayed than what we observe – by at least 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.

These 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). Our experiment controlled which competitors were 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. However, this also raises an interesting question: To what extent are these effects a function of general language processing? It could be that participants who are quicker to activate L2 target words are also quicker in their L1 – They may be better at language processing in general, leading to greater L2 proficiency. Future work should examine whether and how the speed of lexical activation correlates across languages, both for balanced bilinguals and for L2 learners.

Second, the effect of proficiency was not just limited to early activation: we also saw correlations at later timepoints in competition. There was a significant correlation of LexTALE-Esp score on the asymptote of targets (maximum looks; Figure 5B–D) as well as cohorts (offset baseline; Figure 6B and 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 did not resolve competition as fully (and indeed, in the Spanish-Spanish condition as a group they were not able to fully resolve it). 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 organised.

Importantly, differences in these later timepoints are consistent with the pattern of individual differences in L1 ability, seen in kids with development language disorder (McMurray et al., 2010; McMurray et al., 2014). 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-organised L2 lexicon in these adult learners due to individual differences in language ability.

Third, one null result is notable: 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 6(A) shows that there may be some differences in fixations of the cohort between high performers and low performers, however, these differences were not significant. Visually, this pattern 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 L2 cohort due to faster processing. This null result may derive from either – or both – of two causes. It may be that the L2 learners we tested were "too young" in the developmental timeline of their L2 to show this effect, compared to the balanced bilinguals in Blumenfeld and Marian (2013). Second, measuring cohort peak specifically is psychometrically challenging. In a test/retest reliability study of the VWP, 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 & McMurray, 2013), and Monte Carlo simulations suggest that given reasonable models of fixations, cohorts in general may be poorer psychometrically (McMurray, submitted). Therefore, this may have made it more difficult to detect a significant effect in the present study or with the present stimuli.

Across all of these results together, proficiency correlates with several components of lexical access, 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 aspects of word recognition can be described as deriving from a set of latent traits, then a larger battery of standardised tests would better capture and differentiate between the contributions of these factors. The ability to correlate these factors with parameters of competition dynamics could elucidate how latent traits may be important, particularly for the later effects that we observe. In this light, it is worth examining L1 ability alongside these measures in L2, to determine if some general language mechanism may play a role. Second, if language experience is the primary driver of the early correlations with proficiency, then it would be useful to study the effect of an immersive experience on spoken word recognition. For example, how would a summer or semester abroad affect the development within-L2 competition? And finally, cross-sectional or longitudinal work over the course of acquisition will be critical for charting language 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 within their L2 and cross-linguistically. Critically, how these competition dynamics play out over time is related to proficiency in ways that implicate roles of both learning and development, as well as robustness of lexical organisation and language ability. It also underscores the importance of developing real-time 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 such skills develop.

Note

1. Note that this choice is not likely to have a substantial effect on the results after 500 msec -the latest fixation that were excluded would have started at 300 msec, and the average fixation duration is about 200 msec. In contrast, most effects in the present study started after 600 msec.

Acknowledgements

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 assistance with many hours of data collection and assessment scoring.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

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.

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*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In *Memory* (pp. 237–313). Elsevier.

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General*, *139*(2), 299–318. https://doi.org/10.1037/a0018366

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*(5), 633–660. https://doi.org/10.1080/01690960601000746

Blumenfeld, H. K., & Marian, V. (2013). Parallel language activation and cognitive control during spoken word recognition in bilinguals. *Journal of Cognitive Psychology*, *25*(5), 547–567. https://doi.org/10.1080/20445911.2013.812093

Boersma, P. (2006). Praat: doing phonetics by computer. http://www.praat.org/

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition*, *28*(1), 143–153. https://doi.org/10.3758/BF03211582

Chen, H.-C., & Ho, C. (1986). Development of Stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(3), 397–401. https://doi.org/10.1037/0278-7393.12.3.397

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*(4), 335–359. https://doi.org/10.1016/S0022-5371(73)80014-3

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

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001a). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317–367. https://doi.org/10.1006/cogp.2001.0750

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*(5-6), 507–534. https://doi.org/10.1080/01690960143000074

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first-and second-language word recognition: A lexical entrenchment account. *The Quarterly Journal of Experimental Psychology*, *66*(5), 843–863. https://doi.org/10.1080/17470218.2012.720994

Dollaghan, C. (1998). Spoken word recognition in children with and without specific language impairment. *Applied Psycholinguistics*, *19*(2), 193–207. https://doi.org/10.1017/S0142716400010031

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: JSLHR*, *56*(4). https://doi.org/10.1044/1092-4388(2012/12-0145)

Fernandes, T., Kolinsky, R., & Ventura, P. (2009). The metamorphosis of the statistical segmentation output: Lexicalization during artificial language learning. *Cognition*, *112*(3), 349–366. https://doi.org/10.1016/j.cognition.2009.05.002

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*(5-6), 583–607. https://doi.org/10.1080/01690960143000146

Hannagan, T., Magnuson, J. S., & Grainger, J. (2013). Spoken word recognition without a TRACE. *Frontiers in Psychology*, *4*, 563. https://doi.org/10.3389/fpsyg.2013.00563

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*(1), 49–66.

Jonides, J., & Nee, D. E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience*, *139*(1), 181–193. https://doi.org/10.1016/j.neuroscience.2005.06.042

Ju, M., & Luce, P. A. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science*, *15*(5), 314–318. https://doi.org/10.1111/j.0956-7976.2004.00675.x

Kapnoula, E. C., & McMurray, B. (2016). Newly learned word forms are abstract and integrated immediately after acquisition. *Psychonomic Bulletin & Review*, *23*(2), 491–499. https://doi.org/10.3758/s13423-015-0897-1

Kapnoula, E. C., Packard, S., Gupta, P., & McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition*, *134*, 85–99. https://doi.org/10.1016/j.cognition.2014.09.007

Lee, I., & Preacher, K. (2016). Calculation for the test of the difference between two dependent correlations with one variable in common. Retrieved September from http://quantpsy.org

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 608–622. https://doi.org/10.1037/a0029243

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36. https://doi.org/10.1097/00003446-199802000-00001

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*(1), 133–156. https://doi.org/10.1080/03640210709336987

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

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

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

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 576–585. https://doi.org/10.1037/0096-1523.15.3.576

Mcclelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

Mccloskey, M., & Cohen, N. J.. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of Learning and Motivation* (Vol. 24, pp. 109–165). Academic Press. https://doi.org/10.1016/S0079-7421(08)60536-8

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*(8), 1472–1491. https://doi.org/10.1037/dev0000542

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*(4), 961–980. https://doi.org/10.1097/AUD.0000000000000681

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. https://doi.org/10.1016/j.cognition.2019.06.012

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*(4), 1344–1362. https://doi.org/10.1044/2014_JSLHR-L-13-0196

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), 1–39. https://doi.org/10.1016/j.cogpsych.2009.06.003

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. 10 K*. Eurocentres.

Muysken, P. (2013). Two linguistic systems in contact: Grammar, phonology, and lexicon. In *The handbook of bilingualism and multilingualism* (pp. 193–216). John Wiley & Sons.

Perfetti, C. A. (2002). The lexical quality hypothesis. *Precursors of Functional Literacy*, *11*, 67–86.

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*(2), 295–301. https://doi.org/10.1016/S0022-5371(69)80079-4

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*(8), 1325–1332. https://doi.org/10.1037/xlm0000520

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*(12), 1690–1703. https://doi.org/10.1037/dev0000044

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*(2), 314–330. https://doi.org/10.1037/edu0000279

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, *24*(3), 355–367. https://doi.org/10.1007/s10648-012-9201-3

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*(1), 20–45. https://doi.org/10.1016/j.jecp.2007.04.005

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). From the cover: Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*(29), 10393–10398. https://doi.org/10.1073/pnas.0503903102

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, *10*(3), 281–284. https://doi.org/10.1111/1467-9280.00151

Spivey, M., & Mirman, D. (2001). Retroactive interference in neural networks and in humans: The effect of pattern-based learning. *Connection Science*, *13*(3), 257–275. https://doi.org/10.1080/09540090110091830

S. R. E. Builder. (2011). *SR research experiment builder 1.10.165*. SR Research Ltd.

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*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Tzelgov, J., Henik, A., & Leiser, D. (1990). Controlling stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(5), 760–771. https://doi.org/10.1037/0278-7393.16.5.760

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. https://doi.org/10.3389/fpsyg.2016.01120

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*(5), 1011–1032. https://doi.org/10.1017/S0142716418000103

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

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*(3), 495–513. https://doi.org/10.1037/0278-7393.34.3.495

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. https://doi.org/10.1016/j.jml.2018.01.002

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. https://doi.org/10.1080/00221309.2020.1771675