The use of lexical and sublexical cues in speech segmentation by second language learners of English

Candise Yue Lin[1,2] and Min Wang[1,2]
[1]University of Southern California | [2]University of Maryland

This study examined the use of lexical and sublexical cues in speech segmentation by Mandarin L2 learners of English, focusing on two types of lexical cue, lexical knowledge and semantic relatedness, and three coda (sublexical) cues, /n, s, ŋ/ due to their varying phonotactic probabilities in Mandarin and English. Thirty-five native English speakers and 30 L2 learners participated in two experiments. Experiment 1 showed that learners were able to use lexicality as a cue to segment L2 speech. The lexicality effect significantly interacted with L2 proficiency. Experiment 2 showed that learners did not use semantic cues to the same extent as native listeners did. All participants experienced more difficulty with word boundary identification preceded by /s/. This difficulty may stem from weak allophonic cues of /s/ in English. L2 learners with better proficiency may be better at recognizing familiar words from continuous speech, thus more efficiently utilizing the lexicality cue.

Keywords: speech segmentation, lexical knowledge, semantic relatedness, Mandarin, second language acquisition

Introduction

When reading texts, locating the beginning and end of a word is simple because there is a visual gap between each word. When listening to speech, however, there is no similarly reliable cue to indicate word boundaries. Speech signals often do not contain breaks at word edges. Even when breaks occur, they do not coincide with perceived word boundaries (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Nevertheless, there is a vast body of evidence suggesting that listeners could utilize sub-lexical (such as phonotactics, acoustic features such as duration, aspiration, and stress) and lexical cues (such as knowledge about word

https://doi.org/10.1075/jsls.17017.lin
Journal of Second Language Studies 1:1 (2018), pp. 166–198. issn 2542-3835 | e-issn 2542-3843

or sentential context) to facilitate segmentation (e.g., Cutler & Norris, 1988; Mattys, 2004; Mattys, White, & Melhorn, 2005; Newman, Sawusch, & Wunnenberg, 2011; Norris, McQueen, Cutler, & Butterfield, 1997). Mattys et al. (2005) proposed the Hierarchical Framework to capture the weighted importance of these cues. At Tier I of the hierarchy is sentence context which includes semantic, syntactic, and pragmatic cues. Imagine a native English listener trying to recognize the target word "cremate" from the sentence "an alternative to traditional burial is to cremate the dead;" although "mate" is a real word, the listener would not predict a word boundary before "mate" because the sentence context is about burial rather than friendship. Also belonging to Tier I in the hierarchy is lexical knowledge. Words that are familiar to the listener can be segmented and identified simply by matching the sound patterns in the signal with the established phonological representation in the lexicon (e.g., match by car with /baɪ/ and /kɑː/).

If less than optimal listening condition precludes the use of lexical cues, speakers can rely on segmental cues (including phonotactics and acoustics-phonetics) which are at Tier II of the hierarchy (Mattys et al., 2005). For example, it has been found that English speakers are more likely to lengthen word-final syllables (Umeda, 1975; Beckman & Edwards, 1987). Speakers who are sensitive to the acoustic cue of duration may predict a word boundary following the lengthened syllable. Finally, at Tier III of the hierarchy are prosodic cues such as word stress, which are only utilized by native English listeners when lexical or segmental information is masked by noise (Mattys et al., 2005). This is because the location of stress is generally unpredictable in English as stress can fall on any syllable depending on syllable weight and word class. In contrast, stress may be a more useful segmentation cue in languages with demarcative stress such as Hungarian or Finnish in which stress always falls on the word-initial syllable.

The Hierarchical Framework is constructed based on findings from native English listeners. It is not clear whether similar weightings of the various segmentation cues can be generalized to nonnative listeners who learn English as a second language (L2). Since learners may need to establish a relatively large L2 lexicon in order to utilize lexical knowledge, it is likely that L2 learners could not utilize lexical cues to the same extent as native listeners. L2 learners have also been shown to be less sensitive to pragmatic and syntactic structures in L2 sentences (Bardovi-Harlig & Dornyei, 1998; Jiang, 2007). Furthermore, there is evidence suggesting that L1 phonotactic, acoustic, and prosodic structures have cross-linguistic influence on L2 segmentation (Altenberg, 2005; Cutler, 2000; Weber, 2000). Thus, it is possible that nonnative listeners may rely more on sublexical cues over lexical cues.

The current study focused on the comparison between the use of lexical knowledge and phonotactic constraints in Experiment 1 and the use of semantic relatedness and phonotactic cues in Experiment 2. In the context of the current study, lexical knowledge is operationalized as the extraction of familiar words from continuous speech (i.e. "calculus male" should be segmented faster than "baltuluf male" since "calculus" is a familiar word). Semantic relatedness is operationalized as faster identification of word boundary strapping two words that are closely related in meaning compared to another pair that is semantically unrelated. For example, listeners may be faster to identify the word *plant* when it is preceded by *cactus* compared to when it is preceded by *purchase*, even though both words end with /əs/ and /sp/ is a possible consonant cluster onset for the following word. It is important to differentiate between lexical and semantic constraints. While lexical knowledge involves the identification of a known word in the input regardless of its meaning, semantic relatedness takes into account the semantic relevance of a word in a given context (e.g., Blank & Foss, 1978; Tyler & Wessels, 1983). In addition, phonotactic probability is defined as the likelihood of occurrence of a phonological segment or a sequence of phonological segments in a certain position in the word in a given language (Vivevitch & Luce, 2004). When a native English listener hears the phoneme /f/ (as in *knife)* followed by /m/ (as in *man*), he or she can infer that there is likely a boundary between /f/ and /m/ since /fm/ is not a legal consonant cluster at the onset position in English.

The use of lexical cues in native segmentation

Computation models of word recognition such as TRACE (McClelland & Elman, 1986; Rumelhart, McClelland, & the PDP Research Group, 1986) or Shortlist (Norris, 1994) posit that segmentation is the product of word recognition. Lexically driven segmentation is achieved when competition between candidates settles on an acceptable parsing solution that leaves no fragments unaccounted for (Mattys et al., 2005). Building upon these models of word recognition Norris and colleagues (Norris, McQueen, Cutler, & Butterfield, 1997; Norris, McQueen, Cutler, Butterfield, & Kearns, 2001) proposed a lexically based segmentation strategy called the Possible-Word Constraint (PWC). The PWC disfavors interpretations that leave a residue of the input which cannot be identified as one or more words. In a real word-spotting task (with no visual presentation of the target word), native English listeners found it more difficult to spot the real word *apple* in *fapple* compared to in *vuffapple*. Even though both *f* and *vuff* are not members of the English vocabulary, a single consonant can never be a viable candidate word in the English language. While the PWC disfavors sound sequences that cannot be words in the languages, listeners also use real words as a constraint. In a word identification task in which participants saw the target word on the screen and determine whether a following auditory phrase contains the target word, Mattys et al. (2005, Experiment 3) showed that recognition of the target word (i.e. *already*) was faster when it was preceded by a real word context (i.e. *animal already*) than when it was preceded by a nonword context (i.e. *erromal already*) even though both contexts were matched for phonotactic probabilities of the phonemes.

Beyond simply extracting a known word from continuous speech, listeners also consider the semantic relevance in a given context. Mattys et al. (2005, Experiment 5) found that response latency was faster when the target words were semantically related to the preceding context (e.g., *dressing gown* vs. *mayhem gown*). In another study, Dilley, Mattys, and Vinke (2010) asked participants to identify the final word in auditory phrases. The final syllable can be parsed as either a disyllabic or a monosyllabic word (e.g., *turnip* or *nip*). The phrases were either semantically related to the monosyllabic parsing (e.g., *puppy biting cry sister nip*) or to the disyllabic parsing (e.g., *garden veggie crisis turnip*). Participants identified more disyllabic words when the semantic context was consistent with the disyllabic parsing. Consistent with the lexical approach to segmentation, previous research has shown that identification of known words from continuous input and favoring those words most likely given a particular semantic context are both efficient segmentation strategies for native English listeners.

The use of phonotactic cues in native segmentation

English also has a set of phonotactic constraints that can help a listener identify word boundaries. For example, the phoneme /h/ is always syllable-initial and /ŋ/ is always syllable-final (Church 1983). Native English speakers may predict a word boundary preceding /h/ and a boundary following /ŋ/. Phonotactics can occur at the level of a single sound (e.g., /ŋ/ is only allowed syllable-finally) or at the level of biphones. For example, since no /tl/ cluster is allowed within a syllable, English listeners may predict a word boundary between these two phonemes. It is important to differentiate between phonotactic constraints that involve absolute legality (e.g., /h/ can only occur syllable-initially) and those that involve probabilities. For example, English syllables are more likely to end with tense vowels (e.g., /i/) than with lax vowels (e.g., /ʌ/) as lax vowels tend to pull in the following consonant perceptually. Hence, recognition of the word *apple* may be faster in the sequence *vuff-apple* (/vʌf/-apple) than in *veef-apple* (/vif/-apple) as listeners are more likely to perceive /f/ as the coda of the preceding syllable in the case of *vuff-apple*. However, previous studies (e.g., Newman et al., 2011; Norris, McQueen, Cutler, Butterfield, & Kearns, 2001) did not find significant differences in word-spotting latency between *vuff-apple* and *veef-apple*. Moreover, listeners were faster to spot embedded words in sequences when the speaker produced the sequence with the syllabic boundary after the consonant (i.e. *vuff-apple*) compared to sequences when the speaker produced the sequence with the syllabic boundary before the consonant (i.e. *vuh-fapple*). These results suggest that English listeners are less likely to take into consideration the probabilistic phonotactics of vowels in segmentation when there are clear juncture cues present.

Other studies have shown that English listeners were sensitive to the phonotactic probability of consonants at word edges (Vitevitch & Luce, 1998; 1999). Listeners were faster to respond to /sʌv/, in which /s/ has high likelihood of occurrence as an onset and /v/ has high likelihood of occurrence as a coda, than /jʌʃ/, in which /j/ and /ʃ/ have low likelihood of occurrence at their respective positions. The current study examined the use of phonotactic cues in native and nonnative segmentation via the probabilities of English consonant coda.

The use of phonotactic cues in the segmentation of nonnative speech

Nonnative listeners may show different patterns of cue uses in speech segmentation compared to native listeners for at least two reasons. First, L2 learners may not be able to use lexical and semantic cues efficiently until they have developed a decent sized L2 lexicon and constructed semantic representations for L2 words. Thus, semantic and lexical cues may be dependent on the L2 learners' lexical proficiency. Second, there is a high degree of language-specificity for phonotactic cues as certain phonemes and rules exist in one but not another language. As a result the phonotactic constraints in L1 may have cross-linguistic influences on the process of breaking up continuous speech in L2 in light of the typological differences in phonological structures between L1 and L2 (see Cutler, 2000 for a review).

Specifically, Weber (2000) observed that it was easier for native English listeners to detect *luck* in the nonword *moyshluck* than in *moysluck* since /ʃl/ is not a legal sequence in English whereas /sl/ is (e.g., *slack*). The opposite result was observed in highly competent German L2 learners of English. Since in German /ʃl/ is a possible syllable onset but /sl/ is not, L2 learners were more likely to predict a word boundary straddling /sl/. However, there is also evidence showing that L2-specific phonotactic constraints can be learned by L2 learners and used in segmentation of nonnative speech if there is no conflict in the specific phonotactic probabilities between L1 and L2. For example, English L2 learners of French have been shown to be sensitive to the distributional probabilities of French liaison (Tremblay & Spinelli, 2013; 2014). Liaison is a process in which a normally silent word-final consonant is pronounced when it is followed by a vowel-initial word (e.g., *-s* in *sans* [z]*elle* 'without her' vs. *sans peur* 'without fear' in which [z] represents the pronounced liaison consonant). L2 learners of French showed native-like sensitivity to liaison consonants in segmentation even though liaison does not exist in English. Furthermore, Tremblay and Spinelli (2014) found that L2 learner's proficiency in French has little influence on their use of distributional cues.

Weber and Cutler (2006) found that highly proficient German L2 learners of English identified the target word *lecture* faster in the nonword sequence *thrarshlecture* than in the sequence *gorklecture*. While /ʃ l/ is not a legal onset in English and thus signaled a boundary, both /ʃ l/ and /kl/ are legal onsets in German. The findings suggest that advanced L2 learners can exploit phonotactic constraints specific to L2 in nonnative segmentation. Similar results have been replicated with native Arabic L2 learners of English (Al-jasser, 2008). Arabic L2 learners were divided into control and experimental groups in which the experimental group received additional training in English phonotactics for eight weeks. Comparing pre- and post-test performance, the experimental group showed significant improvement in the word-spotting task in the English Boundary condition (i.e. spotting the target word *line* in the sequence *veedline* in which /dl/ is a legal onset in Arabic but illegal in English). Even at pre-test, L2 learners were faster to spot the target word *line* in the Arabic Boundary condition (i.e. in the sequence *veebline* in which /bl/ is an illegal onset in Arabic but legal in English) compared to the No Boundary condition (i.e. in sequence *veefline* in which /fl/ is a possible onset in both languages) whereas the native English listeners did not show a significant difference in word-spotting latency between the Arabic boundary and No Boundary condition. Overall, these results suggest that L2 learners transfer L1 phonotactic constraints when segmenting L2 speech. More importantly, L2 phonotactic constraints can be learned and exploited by nonnative listeners after a short period of training.

The use of lexical and semantic cues in the segmentation of nonnative speech

Only a few of studies have examined the use of cues at the lexical level by nonnative listeners in segmentation of continuous speech (e.g., Sanders, Neville, & Woldroff, 2002; White, Melhorn & Mattys (2010). Hanulikova, Mitterer, and McQueen (2011) tested the PWC with native Slovak speakers who were L2 learners of German. Slovak, West Slavic language, allows words consisting only of a single consonant whereas German does not. These L2 learners found it harder to recognize the target German word *rose* in the sequence *trose* than in *krose* since *k*, not *t*, can be a real word in Slovak. This result demonstrates the effect of L1 lexicality on the segmentation of nonnative speech.

Given the lack of theoretical models for L2 speech segmentation and segmentation in general can be described as a product of lexical access (Mattys et al., 2005; Gow & Gordon, 1995; Norris et al., 1995; 1997), we drew upon models of word recognition in the bilingual literature. A well-known model in bilingual lexical processing, the revised hierarchical model (RHM, Kroll & Stewart, 1994 ; Kroll, van Hell, Tokowicz, & Green, 2010) proposes that the way the bilingual speaker's two languages are linked together is influenced by L2 proficiency. For beginning learners, the link between the lexical representation of L2 words and the corresponding conceptual representation is weaker compared to that between L1 words and concepts. Lexical access of L2 words may require mediation via the L1 translation equivalent until learners have acquired sufficient skill in the L2 to access meaning directly. Although the RHM does not hypothesize about segmentation of L2 speech, based on the predictions that L2 learners have weaker links between L2 words and concepts, it is reasonable to infer that nonnative listeners may not be able to utilize semantic cues efficiently in L2 segmentation compared to native listeners.

On the other hand, previous research with bilingual speakers showed that the amount of lexical competition in spoken word recognition is greater for L2 listeners due to the activation of possible word candidates in the L1 even in a monolingual task done solely in the L2 (Spivey & Marian, 1999; Weber & Cutler, 2004). Such nonselective access was also observed in Mandarin-English bilingual speakers, despite the typological differences in the phonological and orthographic systems between Chinese and English (Zhou, Chen, Yang, & Dunlap, 2010). If nonnative listeners have a larger pool of word candidates (from both L1 and L2) to select from, then the extraction of familiar words in the continuous input as a segmentation strategy may not be as efficient as it would have been for native listeners.

To the best of our knowledge, the study by White, Melhorn and Mattys (2010) was one of the first to examine the use of lexical knowledge in the segmentation of L2 speech by adult learners. In this study, native Hungarian speakers who have achieved various proficiency levels in English completed a cross-modal priming lexical decision task. In each trial, participants were asked to listen to a five-syllable phrase (e.g., *anythingcorri* or *imoshingcorri*) with visual presentation of a three-syllable letter string (e.g., *corridor*) 100ms after the offset of the auditory prime. The critical manipulation was the lexicality of the context word (e.g., *anything* or *imoshing*). Both native English listeners and Hungarian L2 learners responded faster to target words preceded by a real word context and these was no significant correlation between the magnitude of the lexical priming effect and L2 learners' English proficiency. These results suggest that nonnative listeners were able to utilize lexicality as a segmentation cue for L2 speech, despite having a larger pool of possible word candidates to choose from both L1 and L2. Moreover, post-hoc analysis revealed that some participants at the lowest level of L2 competency took substantially longer to respond. After the removal of data from these careful responders, the lexicality effect was no longer significant in this group of L2 learners with the lowest L2 proficiency. It appears that the lexical segmentation strategy via the recognition of known words do require certain degree of familiarity with the English vocabulary.

Even though studies of bilingual lexical processing (e.g., Zhou et al., 2010) and White et al. (2010) seem to suggest inconsistent findings, it should be noted that these studies involve different research questions and experimental designs. While Zhou et al. used a masked visual priming paradigm in which both the prime and target were individual words, White et al. employed a cross-modal priming paradigm in which the auditory prime was a phrase. Since White et al.'s participants were Hungarian L2 learners of English; it was not clear whether these results could be generalized to English learners with other L1 backgrounds. The current study aimed to extend these findings to a L1-L2 combination that involves a non-Indo-European language and a different task paradigm.

The current study

The goal of the current study was to compare the use of lexical and phonotactic cues between native English listeners and Mandarin L2 learners of English. This L2 group was chosen because of the limited coda inventory and simpler syllabic structure in Mandarin that form an obvious contrast to English. Mandarin has only two legal coda consonants, /n/ and /ŋ/, and there is no consonant cluster. The maximal syllable structure is CGVX where C is an onset, G is a glide, V is a vowel, and X is a coda. In contrast, English has a more expansive coda inventory and the maximal syllable structure can contain up to three onset consonants, a vowel, and five coda consonants. These cross-linguistic differences allowed us to examine the influence of L1 phonotactic cues on L2 segmentation.

The current study focused on three consonants, /n/, /ŋ/, and /s/. Regarding absolute legality, /n/ can occur in both word-initial and word-final positions in both English and Mandarin. In contrast, /ŋ/ can only occur syllable-finally in both languages. While /s/ can occur in the word-initial position in both languages, it is not allowed syllable-finally in Mandarin. To examine the phonotactic probabilities of these phonemes, probability and frequency of occurrences were obtained from SUBTLEXus (derived from American English subtitles, total = 51 million words, Brysbaert & New, 2009) and polymorphemic words such as *knows*, *beginning*, and *organization* were included in this analysis. Out of the 74286 unique words represented in the corpus (see Table 6 in Supplementary Materials), there are 3.5 times as many /n/-final words as /n/-initial words and there are 2.4 times as many /s/-final words as there are /s/-initial words. In addition, /n/-final words are 2.6 times more likely to be heard than /n/-initial words whereas /s/-final words were 1.8 times more frequent than /s/-initial words. These corpus data indicate that although in general /s/ has a higher probability of occurrence and higher frequency than /n/, both phonemes are more likely to occur in the word-final than word-initial position. The ratio between word-initial and word-final positions is similar for the two phonemes in terms of likelihood of occurrence (3.5 times for /n/ and 2.4 times for /s/) and word frequency (2.6 times for /n/ and 1.8 times for /s/). A listener who is sensitive to these distributional cues should be able to identify a word boundary following /s/ as fast as they identify a word boundary following /n/.

Probability and frequency of occurrence of the target phonemes in Mandarin were also obtained from SUBTLEX-CH, a corpus based on Chinese Film subtitles (Cai & Brysbaert, 2010, total = 33.5 million words or 46.8 million characters). Out of the 99121 unique words (including morphologically complex words) represented in the database, there are 9.1 times as many /n/-final words as /n/-initial words in the corpus. Moreover, /n/-final words are 1.8 times more likely to occur than /n/-initial words. Therefore, Mandarin and English are similar in terms of the distributional cues for /n/ in terms of more word-finals than word-initials. If Mandarin L2 learners are sensitive to the distributional cues in L2 English or if the L2 learners tend to utilize L1 distributional cues to segment L2 speech, their response pattern for /n/-final stimuli should be similar to that of native listeners. However, in the case of /s/-final words, these two possibilities can be teased apart when examining L2 learners' segmentation. If L2 learners are sensitive to the distributional cues in English, they should be able to segment /s/-final words as fast as /n/-final words. However, if L1 phonotactic constraints have cross-linguistic influence on nonnative listeners' segmentation, they may be slower to identify a word boundary following coda /s/ than coda /n/. Finally, although /n/ is more likely to occur word-finally than word-initially in both languages, /ŋ/ is not allowed in the syllable-initial position. Therefore, both language groups of listeners should be faster to identify a word boundary following coda /ŋ/ than coda /n/.

In summary, the current study examined two types of lexical cues (lexical knowledge and semantic relatedness) and phonotactic constraints (focusing on the three coda cues). English proficiency was operationalized as participants' performance on an objective assessment of competence in English syntax, semantics, and vocabulary. Since L2 learners have lower proficiency and weaker links between lexical and conceptual representation of L2 words than native listeners, we predicted that they should utilize both types of lexical cues to a lesser extent than native listeners in English speech segmentation. For L2 learners' use of phonotactic cues, we predicted that segmentation of the target words would be faster when the coda of the preceding words is consistent with L1 phonotactic constraint (i.e. /ŋ/ for L2 learners) than when the coda violates L1 phonotactic constraint (i.e. /s/ for L2 learners).

Experiment 1: Phonotactic cues versus lexical knowledge

The use of phonotactic cues and lexical knowledge was examined in a word identification task. In English, both /n/ and /s/ are more likely to occur in word-final than word-initial positions and both /n/-final and /s/-final words have higher frequency than /n/-initial and /s/-initial words. Hence, it was hypothesized that native listeners would segment auditory stimuli with coda /n/ as fast as they segment those with coda /s/. Since word-initial /ŋ/ violates the phonotactic constraints in English but word-initial /n/ or /s/ does not, native listeners' word identification was expected to be faster for coda /ŋ/ than for /n/ or /s/. In Mandarin, word-final /n/ has a higher likelihood of occurrence and higher word frequency than word-initial /n/ and /ŋ/ is not allowed syllable-initially. L2 learners' word identification with /n/ or /ŋ/ was hypothesized to be similar to that of native listeners'. Since word-final /s/ is not allowed in Mandarin, we predicted that L2 learners would take longer to identify target words following /s/ if L1 influence occurs. For the use of lexical knowledge, native listeners would respond faster to the target word preceded by a real word context than by a nonword context. In contrast, L2 learners may not show consistent use of lexical knowledge due to their lower proficiency in English.

Participants

There were two groups of participants, a monolingual English group ($N=35$, male$=9$) and a L2 learner group with L1 Mandarin ($N=30$, male$=11$). At the time of testing, all participants were undergraduate or graduate students in a mid-Atlantic university in the United States. All participants completed the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian, Blumenfeld, & Kaushanskaya, 2007). In this questionnaire, participants provided information about age of acquisition (AOA), length of U.S. residence, and amount of language use in L1 and L2. Participants also self-rated their speaking, listening, and reading abilities in L1 and L2. An objective proficiency measure, the cloze test (Bachman, 1982), was included. It is a fill-in-the-blank activity that assesses English syntactic and lexical proficiency. This test was not timed and the total possible score is 50. Table 1 shows the demographics of the L1 and L2 groups based on the LEAP-Q and the cloze test scores. None of the English speakers had any experience with Mandarin. All of them had learned a foreign language in high school to fulfill graduation requirement. However, none of them speak that foreign language fluent enough to consider it as their second language and chose not to report this in the LEAP-Q. Since there could be cultural differences in self-rating, we included cloze test scores as a fixed effect in all subsequent analyses to account for the effect of proficiency in cue use.

Table 1. Demographics of the two language groups

	English (N=35)	**Mandarin** (N=30)
Chronological age	20.0 (2.12)	24.4 (2.69)
Cloze test (out of 50)	46.3 (2.25)	37.8 (3.69)
Self-rated proficiency for understanding spoken language in English (out of 10)	9.73 (.449)	7.07 (1.41)
Age of L2 acquisition	N/A	9.6 (2.97)
Length of U.S. residence (years)	N/A	1.39 (1.37)
Percentage of L2 use (out of 100)	N/A	42.9 (12.6)

Note. Standard deviations in parenthesis.

Materials and design

This experiment employed a word identification task (adapted from Mattys et al., 2005; Mattys & Melhorn, 2007). In each trial, a visual target (e.g., *already*) appears on the screen for 1000ms, followed by the immediate presentation of an auditory phrase. Participants were asked to decide whether the auditory phrase contains the target word they have seen before. The auditory phrase consists of six syllables in which the first three syllables constitute the *auditory context* and the last three syllables make up the *auditory target* (e.g., *everyone-already*). The design of this experiment is 2×3 factorial with the two factors being lexicality of the contexts (real words or nonwords) and syllable codas of the contexts (/n, ŋ, or s/).

An initial list of 90 trisyllabic words was generated via the English Lexicon Project (Balota et al., 2007) based on their matched written frequency from the HAL database (Lund & Burgess, 1996). Thirty of the words have /n/ as a coda, thirty ends with /ŋ/ and thirty with /s/. This list was sent to 10 Mandarin L2 learners of English in the same population where the current sample was drawn from.

The raters rated how familiar they were to each word based on a 7-point Likert scale with 1 being "not familiar at all" and 7 being "very familiar". Only words with a mean familiarity rating higher than 6.5 were selected. For words in the coda /s/ condition, it was also ensured that the /s/ is not realized as /z/ word-finally. The final stimuli list consisted of a total of 54 words, 18 in each of the three coda conditions. Half of the 18 words in each condition were initial-stressed while the other half were medial-stressed. Words were matched across the coda conditions on written and spoken frequency, familiarity, number of letters, number of phonemes, the size of phonological neighborhood, and uniqueness point (Table 7 in Supplemental Materials). However, the /ŋ/ condition happened to have significantly higher biphone token frequency and biphone type frequency than the /s/ and /n/ conditions ($F(2,57) = 23.881$, $p < .001$; $F(2,57) = 56.446$, $p < .001$, respectively), probably due to the highly frequent co-occurrence of the biphone /ɪŋ/ as the English present tense marker *-ing*.

Nonwords designed to match the real word contexts were created using the Phonotactic Probability calculator (Vitevitch & Luce, 2004) so that each phoneme in the nonword is matched with the phoneme in the corresponding real word in terms of position-specific probability. Fifty-four nonwords were created, 18 in each of the coda conditions. The nonwords had the same stress pattern as their corresponding real words. Recordings of the nonwords were sent to four native English speakers who were drawn from the same population as the native English-speaking participants in the current study and blind to the hypotheses of the current study. The raters were asked to listen to each sound file and judge how much each nonword sounded like a real word based on a Likert scale of 1–7 with 1 being "does not sound like a real word at all" and 7 being "sounds very much like a real word." One-way ANOVA showed that there was no significant difference in nonword ratings among the three coda conditions ($F(2, 51) = 2.043$, $p = .14$). The means rating was 3.972 in the coda /n/ condition, 3.889 in /ŋ/, and 3.208 in /s/. Visual targets were 18 vowel-initial words to make segmentation more challenging. These target words were selected from an original list of 30 words and 12 words with a familiarity rating lower than 6.5 were excluded.

As a result of the 2×3 factorial design, each target was paired with six different contexts for the auditory phrase (See Appendix A in supplemental materials). Two lists were created so that List 1 contained three phrases from one set and List 2 contained the other three phrases from the same set. There were a total of 54 critical trials in each list. Considering that each participant would see the same target word three times during the experiment, trial order was entered into the linear mixed-effect models as a fixed effect (see Results section).

To prevent participants from developing processing strategies for the specific codas or only focusing on the last three syllables of the auditory phrase, three types of fillers were created, with 60 trials in each type. For the first type of fillers, the visual target matched neither the auditory context nor the auditory target. Half of the auditory contexts were real words and the other halves were nonwords while the auditory targets were all real words. For the second type of fillers, the visual target matched the context in the auditory phrases. Finally, for the third type of fillers, the visual target matched neither the auditory context nor the auditory target. All auditory contexts were real words while half of the auditory targets were nonwords and half of them were real words. There were a total of 234 trials, 54 critical trials and 180 filler trials, with an equal number of positive and negative responses. The trials were pseudo-randomized so that there were at least 70 trials separating the same target word to prevent any repetition priming effect. There were no more than three "Yes" or "No" responses consecutively. One female native speaker of American English (with Northeastern dialect) recorded all stimuli. She pronounced each full phrase without interruption (e.g., everyonealready). Recording was done in a quiet room using an Audio-Technica ATR 20 low impedance microphone. The sounds were recorded using SONY Sound Forge and the files were stored as uncompressed WAV, digitized at 44.1kHz at 16bits. Each stimulus was manually cut using Praat (Boersma & Weenik, 2010) to ensure there was no silence at the onset or offset of the phrase. Intensity was normalized at 70db.

Procedure

Participants were tested individually in a quiet room. They were randomly assigned to List 1 or 2. The experiment was implemented via the E-prime software (Psychology Software Inc., Pittsburgh, PA). E-prime recorded response times (RT) starting at the onset of the auditory stimuli. For statistical analysis, the length of the auditory context words was subtracted from the RTs to ensure that the RTs indeed reflect only the recognition of the auditory target words. For each trial, participants first saw a fixation "+" in the center of the computer screen for 500ms. Then they saw the target word (e.g., *already*) which stays on the screen for 1000ms. Immediately following the visual word, participants heard the auditory phrase (e.g., *everyone-already*) and they were instructed to decide whether the auditory phrase contains the visual target word by pressing the keys labeled "Yes" or "No". Speed and accuracy were emphasized. The inter-trial interval was 1000ms. Participants completed eight practice trials with feedback before the test trials to familiarize with the procedure. There was no feedback during the test trials.

Results and discussion

RT data for incorrect responses were excluded, resulting in the loss of 4.6% of total data. RT data were log-transformed to improve normality. All analyses were carried out in R Studio, an open-source programming environment for statistical computing (R Development Core Team, 2007). Untrimmed and unaveraged RT data (3349 data points) were submitted to linear mixed-effects (LME) modeling using the "lmerTest" package (Kuznetsova, Brockhoff, & Christensen, 2013). Accuracy data (3511 data points) was analyzed with the binomial function. Multiple comparisons of means were conducted for each significant effect and interaction using the "lsmeans" package (Lenth, 2017) with Tukey contrasts and adjusted *p*-values. The anova() function from the "lmerTest" package displays the LME output in an ANOVA table, providing *F*-statistics and corresponding *p*-values of Type III Hypotheses and Satterthwaite approximation for degrees of freedom for the fixed effects. The model was built based on a forward algorithm in which the baseline model was a regression line of log RT or accuracy rates (with the binomial function) with random intercepts for subjects and items. Each fixed effect, including language group (Mandarin vs. English), context lexicality (real words vs. nonwords), coda (/n, ŋ, s/), list, and trial order (hereafter, TO), and interaction terms were individually added to the model and tested by comparing the log likelihood ratio to that of the simpler model. Only effects that significantly improved the model fit were retained. After the fixed effects have been established, a random slope was individually added for each of the significant fixed effects. Random slopes that did not significantly improve the model fit or resulted in a model that failed to converge were removed. We did not include proficiency in the model because correlation between the effects makes the model unstable and the results difficult to interpret. Language group and cloze test scores were highly correlated since the native listeners were significantly more proficient than the L2 learners ($t = -.64.69$, $p < .0001$).

The most parsimonious model with the best-fit to the RT data (Table 3) included fixed effects of language, group, lexicality, coda, list, and TO; a four-way interaction among language, lexicality, list, and TO; four three two-way interactions: (1) language, lexicality, and TO, (2) list, lexicality and TO, (3) List, language, and TO, and (4) list, language and lexicality; six two-way interactions: (1) list and language, (2) list and lexicality, (3) language and lexicality, (4) list and TO, (5) language and TO, and (6) lexicality and TO; a by-subject random effect, a by-item random slope for TO. In light of the significant four-way interaction, we analyzed the data separately for the two lists and first, second, or third appearance of the target words, creating six mixed-effect models with language, lexicality as fixed effects and subject and items as random effects.

Table 2. Descriptive statistics of response times and accuracy in Experiment 1

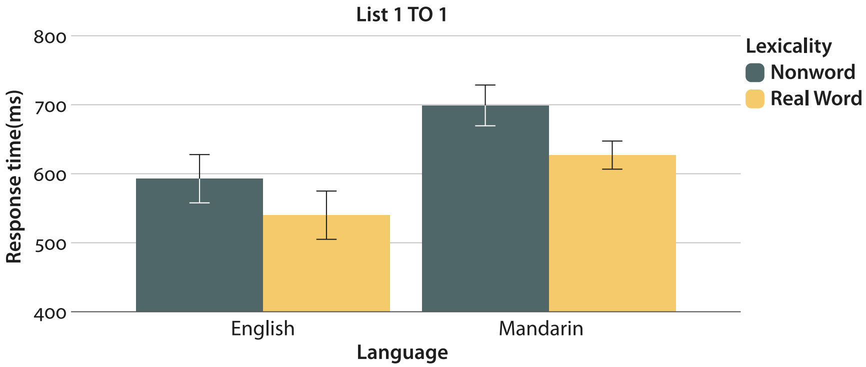
	English				Mandarin			
Lexicality	Nonword		Word		Nonword		Word	
	RT	Acc	RT	Acc	RT	Acc	RT	Acc
Coda /n/	564 (164)	.952 (.214)	667 (228)	.965 (.184)	667 (228)	.968 (.174)	1259 (221)	.992 (.086)
Coda /ŋ/	543 (172)	.949 (.220)	520 (171)	.975 (.157)	677 (183)	.958 (.198)	627 (180)	.970 (.169)
Coda /s/	582 (197)	.896 (.305)	555 (259)	.964 (.184)	696 (199)	.911 (.285)	679 (207)	.940 (.237)

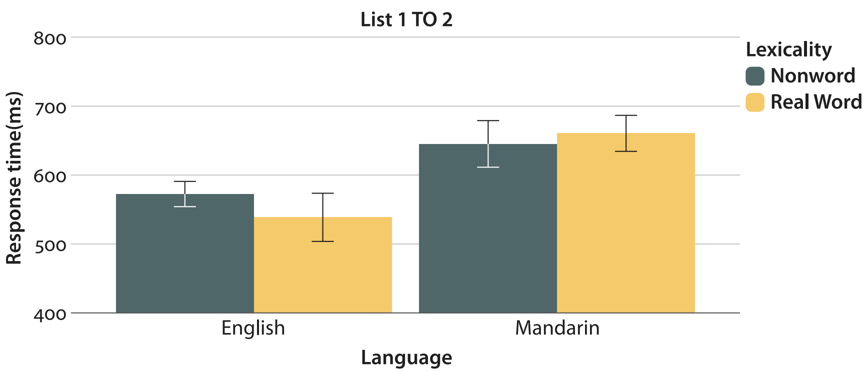
Note. Values in parenthesis represent standard deviations.

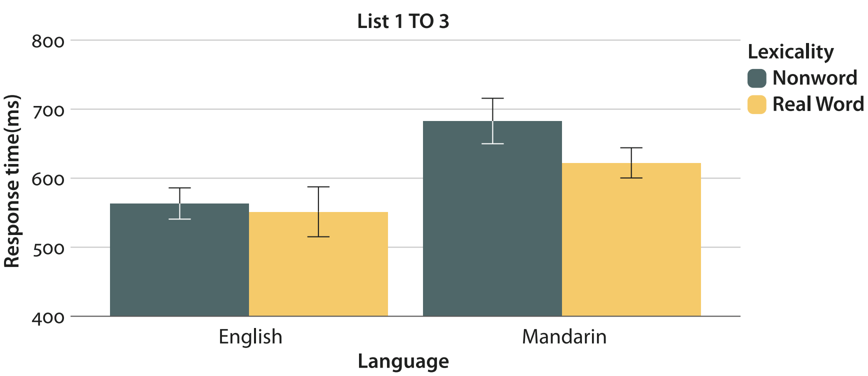
Results (Figure 1) revealed that the two-way interaction between language and lexicality was significant for List 1 and TO2 (F=9.798, p<.001) and for List 2 and TO2 (F=3.954, p=.047) (all other ps>.07). Pairwise comparison showed that for stimuli in List 1 and TO2, although native listeners were significantly faster than L2 learners at identifying target words, the magnitude of this advantage was greater for real word context (t=−4.669, p<.001) than for nonword context (t=−2.489, p=.075). Similarly for stimuli in List 2 and TO2, English listeners showed significantly faster word identification latency than L2 learners, the magnitude of this advantage was greater for real word context (t=−4.533, p<.001) than for nonword context (t=−3.302, p=.010). These results may suggest that L2 learners were less sensitive to the lexical cue compared to their native counterparts.

The most parsimonious model with the best fit for accuracy (Table 3 lower part) included fixed effects for language, coda, and lexicality and random effects of subject and item. The significant effect of lexicality indicated that both language groups were more accurate in word identification when context was a real word than when it was a nonword. For the significant effect of coda, post hoc comparisons showed that both language groups were significantly less accurate identifying word boundary preceded by /s/ compared to /n/ (z=−3.008, p=.007) or /ŋ/ (z=−2.704, p=.019). There was no significant difference in accuracy between /n/ and /ŋ/ (z=−.297, p=.952).

To examine the effect of English proficiency on nonnative listeners' use of lexical knowledge and coda cues, we built another set of mixed-effect models only including the data from the L2 learners with lexicalty, coda cues, proficiency (i.e. cloze test scores as a continuous variable), TO, and the interaction terms between these variables as the fixed effects and random effects of subjects and items. For RT data, the best-fit model did not include an interaction between term lexicality and proficiency while the interaction between coda and proficiency was only marginally significant (F=2.840, p=.058). For the accuracy data,


List 1 TO 1
800
700
600
500
400
Response time(ms)
English
Mandarin
Language
Lexicality
Nonword
Real Word


List 1 TO 2
800
700
600
500
400
Response time(ms)
English
Mandarin
Language
Lexicality
Nonword
Real Word


List 1 TO 3
800
700
600
500
400
Response time(ms)
English
Mandarin
Language
Lexicality
Nonword
Real Word

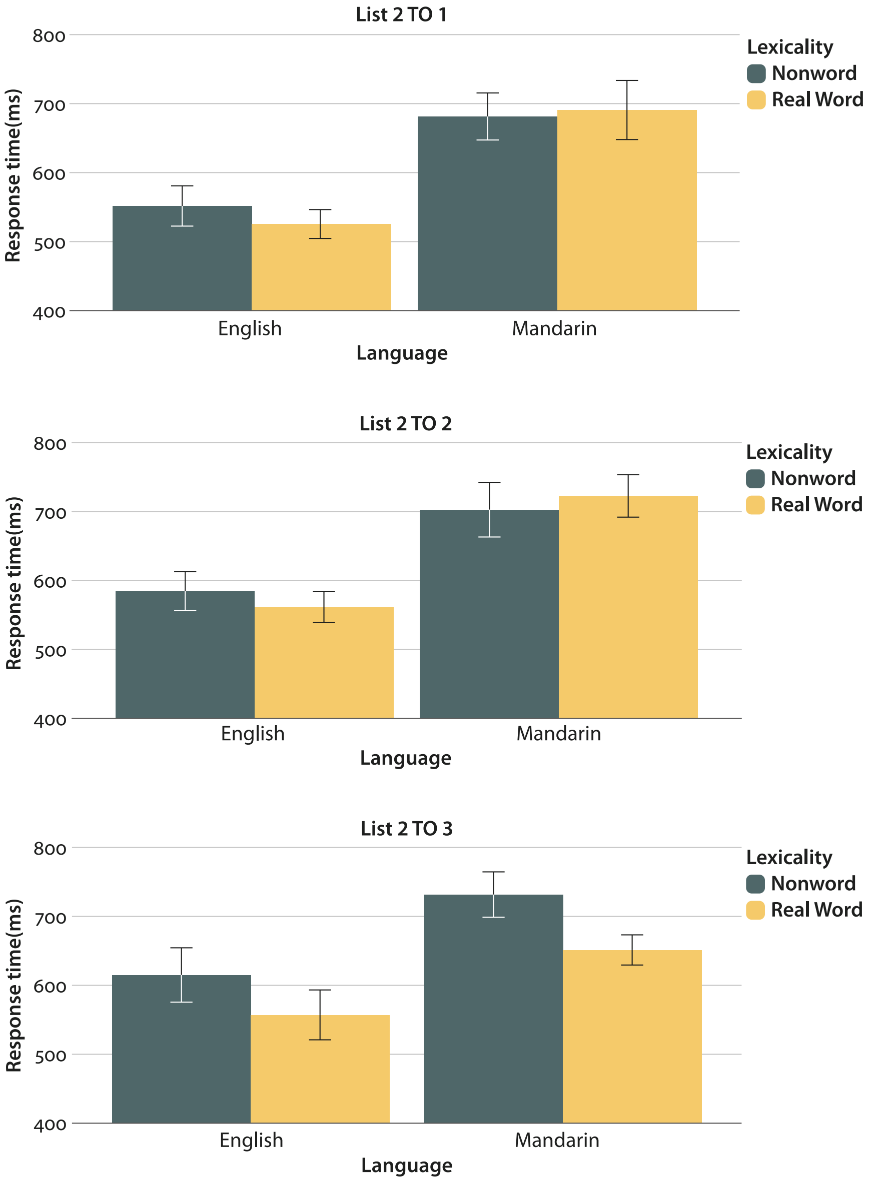


Figure 1. Three-way interaction among language group, context lexicality, and trial order

the best-fit model showed a significant interaction between lexicality and proficiency ($F = 4.541$, $p = .033$). To analyze this significant interaction, L2 learners were divided into two groups of high and low proficiency based on a median-split of cloze test scores. Learners scored 38 or above were categorized as high proficiency

Table 3. Results from the linear mixed-effects model analysis of log RT and accuracy rates in Experiment 1

Log RT

Fixed Effects	**Sum Sq**	**Mean Sq**	**Num DF**	**Den DF**	***F*-value**	**Pr(>F)**
List	.001	.001	1	74.0	.287	.594
Language	.079	.079	1	64.9	29.693	<.001
Lexicality	.044	.044	1	55.6	16.632	<.001
TO	.001	.001	2	43.1	.245	.783
Coda	.044	.022	2	55.1	8.250	<.001
List × Language	.000	.000	1	64.9	.136	.714
List × Lexicality	.014	.014	1	524.5	5.275	.022
Language × Lexicality	.018	.018	1	3548.6	6.662	.010
List × TO	.069	.035	2	703.3	13.067	<.001
TO × Language	.028	.014	2	3548.4	5.327	.005
TO × Lexicality	.000	.000	2	43.2	.0138	.986
List × Language × Lexicality	.007	.007	1	3548.7	2.778	.096
List × Language × TO	.009	.004	2	3548.5	1.686	.185
List × Lexicality × TO	.102	.051	2	733.2	19.286	<.0001
Language × Lexicality × TO	.014	.007	2	3548.2	2.655	.070
List × Language × Lexicality × TO	.016	.008	2	3548.3	3.039	.048

Accuracy

Fixed Effects	**Sum Sq**	**Mean Sq**	**Num DF**	**Den DF**	***F*-value**	**Pr(>F)**
Language	.020	.020	1	64.766	.486	.488
Coda	.456	.228	2	57.229	5.466	.007
Lexicality	.285	.285	1	57.250	6.828	.011

Note. Language = English vs. Mandarin. Coda = /n/, /ŋ/, vs. /s/. Lexicality = real word vs. nonword context. TO = trial order.

(N = 16, M = 40.92) while those scored 37 or below were categorized as low proficiency (N = 14, M = 35.35). Post-hoc comparison showed that for the low proficiency group, there was no significant difference in accuracy irrespective of the context word is a real word or a nonword (z = −.193, p = .847); for the high proficiency group, accuracy was significantly higher for real word context than for nonword context (z = 2.284, p = .022).

Results regarding lexical knowledge were partially consistent with the hypothesis. Findings from both RT and accuracy data indicated an overall lexicality effect for both native and nonnative listeners, suggesting that L2 learners were able utilize the extraction of known words as a segmentation strategy for English. However, the lexicality effect in nonnative listeners was not robust since there were significant interaction with lists and trial orders. With certain sets of stimuli, L2 learners may be less sensitive to the lexicality cue compared to native listeners. Furthermore, the accuracy data showed a significant relationship between the use of lexical knowledge and nonnative listeners' English proficiency. Specifically, L2 learners with more advanced proficiency showed a significant lexicality effect whereas those with lower proficiency did not benefit from the presence of real word context in their identification of the target word. This result is consistent with White et al. (2010) in which Hungarian learners of English with the lowest level of L2 competence level did not show a significant lexicality effect. During the stimuli selection process, we ensured the context words were familiar to the population of Mandarin-speaking L2 learners of English from which the current nonnative group of participants were drawn from, it appears that the segmentation by subtraction of familiar words strategy require a decent level of English proficiency.

For the use of coda cues, findings from accuracy data indicated that both native and nonnative listeners experienced more difficulty segmenting target words preceded by /s/ coda than those preceded by /n/ or /ŋ/. Since both native English listeners and L2 learners show similar patterns of results and there was no significant interaction between language group and coda cues, it is unlikely that the coda /s/ effect for L2 learners results from the influence of L1 phonotactic constraints. Furthermore, the ratio between word-initial and word-final positions is similar for /n/ and /s/ in terms of likelihood of occurrence; therefore, the heightened difficulty to segment words ending with /s/ compared to those ending with /n/ should not be considered as a result of the phonotactic probability in English segmentation.

However, the difficulty with coda /s/ may be explained by the less robust allophonic cues of fricatives (Christie, 1974; Lehiste, 1960; Nakatani & Dukes, 1977). Allophonic variations are defined as the differences in acoustic details of how phonetic segments are pronounced in various syllable positions in fluent speech (Newman et al., 2011). Lehiste (1960) identified fricatives as the consonant class with weak allophonic cues. Only a few potential acoustic cues such as the duration of the fricative and the duration of the preceding vowel can differentiate between syllable-initial and syllable-final fricatives. Newman et al. (2011) found that consonants with stronger allophonic cues have a greater impact on segmentation than consonants with weaker allophonic variations. Thus, even though /s/ is a legal coda in English, its weaker allophonic variations may make it more difficult to use it as a cue to word boundaries compared to /n/. The design of the current experiment did not allow us to tease apart the influence of acoustic-phonetic vs. phonotactic cues on native English segmentation.

Another possible explanation for participants' poorer performance in the /s/ condition is that there were more embedded words and unintentional words in the target and across the word/nonword boundaries in the /s/ condition, making it more likely to assign /s/ to the onset of the following target word. For example, *sofraness-example* may be interpreted as *sex* and *ample* and *igstandous-unable* may be interpreted as *sun* and *able*. However, this may also happen with the /n/ condition in which *protresion-unable* may be interpreted as *nun* and *able* while *protresion -unstable* may be interpreted as *nun* and *stable*. This explanation would be tested in Experiment 2 with different stimuli. If the difficulty with coda /s/ persists, then participants' poor performance in the /s/ condition in the current experiment was unlikely to be stimuli-specific.

Experiment 2: Phonotactic cues versus semantic cues

Experiment 1 showed evidence for the use of lexical knowledge in speech segmentation for both native and nonnative listeners. Experiment 2 examined whether this finding can be extended to a lexical cue at the phrasal level, that is, semantic relatedness between pairs of words. Based on the RHM (Kroll & Stewart, 1994), the use of semantic cues in L2 segmentation may be less efficient if learners have to activate semantic representation of L2 words via L1 translations. Although learners, irrespective of their L2 competence, could develop awareness of what constitutes a real word in English, they may need advanced L2 proficiency to establish rapid and automatized access to semantic representation of L2 words.

Experiments 1 and 2 were similar in terms of using the same codas (e.g., /n, ŋ, s/) and employing the word identification task. There were two main differences, however. First, only real word stimuli were used in Experiment 2. Second, the use of semantic cues in segmentation was examined by manipulating the semantic relatedness of *auditory context* and *auditory target*. Previous research has shown that participants react faster to the target word (i.e. *nurse*) if it is semantically related to the preceding word (i.e. *doctor*) (Perea & Rosa, 2002). The use of semantic relatedness in segmentation was operationalized as faster word identification for target words preceded by semantically related context than those preceded by unrelated context. Based on the RHM, we hypothesized that native English listeners, but not L2 learners, would show a significant semantic effect. Furthermore, based on the results from Experiment 1, we predicted that both native and nonnative listeners would be slower to identify target words preceded by context words with /s/ coda compared to those with /n/ or /ŋ/.

Participants

Participants were the same as those from Experiment 1. Both experiments were administered in the same testing session, 30-minutes apart, in counterbalanced order.

Materials and design

The design was 2×3 factorial with the two factors being the semantic relatedness of the *context* and *target* (related or unrelated) and syllable coda in the *context* (/n, ŋ, s/). In the critical trials, all *auditory targets* were disyllabic while most *auditory contexts* were trisyllabic words (there were several words such as *sizzling*, *struggling*, *champion*, *guardian*, and *Celcius* which can be categorized as disyllabic depending on the regional dialect). Twenty disyllabic high-frequency nouns with concrete meanings were generated from the English Lexicon Project (Balota et al., 2007). For each target word, three semantically related words and three semantically unrelated words were created for each of the three coda conditions. Thus, the original set of stimuli consists of a total of 360 context-target phrases. This list was sent to the L2 raters who completed the familiarity ratings for the previous experiment and an additional 10 native English speakers. They were asked to judge how much is the target word related to the context on a 7-point Likert scale with 1 being "very unrelated" and 7 being "very related". For each coda condition, the context word with the highest relatedness rating was chosen for the related condition and the context word with the lowest relatedness rating was selected for the unrelated condition. This resulted in 18 related and 18 unrelated context words for each coda condition. Context words in the related condition were rated significantly more related to the target words than those in the unrelated condition ($F(118) = 21.12$, $p < .001$). Words in the related and unrelated conditions were matched on all of the relevant properties (see Table 8 in Supplemental Materials). In the coda /n/ condition, 12 words in the related condition were medially stressed while 13 words in the unrelated condition were medially stressed; in the /ŋ/ condition, there were 12 medially-stressed words in both related and unrelated conditions; and in the /s/ condition, there were seven medially-stressed words in both related and unrelated conditions. Due to nonnative listeners' limited vocabulary size and the stringent stimuli selection criteria, we could not match the words' stress location across the three coda positions. Nevertheless, stress location is relatively well matched between related and unrelated conditions.

As a result of the 2×3 factorial design, each target was paired with one semantically related and one semantically unrelated context in each of the three coda conditions, yielding six possible combinations (see Appendix B in Supplementary

Materials). Similar to Experiment 1, two lists were created so that List 1 contained three phrases from one set and List 2 contained the other three phrases from the same set and trial order was added as a fixed effect in subsequent mixed-effects modeling. To prevent participants from developing processing strategies, three types of fillers were created, with 60 trials in each type. For the first type of fillers, the visual targets were disyllabic words that matched neither the auditory context nor the auditory targets; the auditory context and target were semantically related. For the second type of fillers, the visual targets were trisyllabic words that matched the auditory contexts; the auditory contexts and targets were semantically unrelated. Finally, for the third type of fillers, the visual targets were also trisyllabic words but they matched neither the auditory contexts nor the auditory targets; the auditory contexts and targets were semantically unrelated. There were a total of 234 trials, 54 critical trials and 180 filler trials, with an equal number of positive and negative responses. The trials were pseudo-randomized so that there were at least 70 trials separating the same target to prevent any repetition priming effect. There were no more than three "Yes" or "No" responses consecutively.

The stimuli were recorded by the same female native English speaker and cut manually and normalized for intensity as in Experiment 1

Procedure

Procedure was the same as that in Experiment 2. In each trial, participants saw a target word (e.g., *mentor*) on the screen and decided whether the auditory phrase (e.g., *studying-mentor* or *betraying-mentor*) contains this target word. Although the correct response would be "yes" for both examples, participants would respond faster to *studying-mentor* than *betraying-mentor* if semantic relatedness facilitates word segmentation.

Results and discussion

The analytical approach was similar to that in Experiment 1. RT data for incorrect responses were excluded (4.6% of total data). RT were log-transformed to improve normality. The baseline model was a regression line of log RT (3345 data points) or accuracy (3511 data points) with random intercepts for subjects and items. Using the forward algorithm, the following fixed effects and their interaction terms were individually entered: language group (Mandarin vs. English), semantic relatedness (related vs. unrelated), coda (/n, ŋ, s/), list, and trial order (hereafter, TO).

Overall, results (Table 4) showed a significant effect for semantic relatedness ($z = 5.57$, $p < .001$). In addition, both language groups were slower to identify word boundary following coda /s/ and faster to locate word boundary following /ŋ/

compared to /n/. The most parsimonious model with the best fit for RT data (Table 5 upper part) included fixed effects of language, coda, semantic relatedness, TO, and an interaction term between language and TO, a by-subject random effect, a by-item random slope for TO. For the effect of coda, participants were significantly faster to identify word boundaries preceded by /ŋ/ than /n/ ($z = -2.444$, $p = .038$); participants were also significantly slower to segment words preceded by /s/ than /n/ ($z = 6.647$, $p < .001$) or /ŋ/ ($z = 9.121$, $p < .001$). For the significant interaction between language group and TO, pairwise comparison revealed that English listeners were significantly faster than nonnative listeners when the target word appeared for the first time ($z = 2.893$, $p = .033$), when the target word appeared for the second time, the difference in RT between the language groups was only marginally significant ($z = 2.658$, $p = .061$), when the target word appeared for the third time, the difference in RT between the language groups was no longer significant ($z = 1.283$, $p = .746$).

Table 4. Descriptive statistics of response times and accuracy in Experiment 2

	English				**Mandarin**			
Semantic	**Related**		**Unrelated**		**Related**		**Unrelated**	
	RT	**Acc**	**RT**	**Acc**	**RT**	**Acc**	**RT**	**Acc**
Coda /n/	496 (163)	.971 (.167)	529 (270)	.937 (.243)	595 (212)	.969 (.174)	611 (180)	.986 (.119)
Coda /ŋ/	481 (166)	.965 (.183)	507 (161)	.946 (.227)	587 (176)	.971 (.167)	605 (231)	.965 (.182)
Coda /s/	522 (154)	.956 (.206)	563 (175)	.911 (.285)	664 (217)	.931 (.254)	670 (205)	.915 (.280)

Note. Values in parenthesis represent standard deviations.

The best-fit model for accuracy (Table 5 upper part) included fixed effects of language group, coda, semantic relatedness, and an interaction term between language group and semantic relatedness, and random effects of subject and item. For the significant interaction between language group and semantic relatedness, pairwise comparison (Figure 2) showed that native English listeners were significantly more accurate to identify target words preceded by related context compared to those preceded by unrelated context ($z = -3.387$, $p = .004$) whereas this semantic effect was not significant for nonnative listeners ($z = -.166$, $p = .998$). The effect of coda cues indicated that participants were significantly less accurate to identify target words preceded by /s/ than by /n/ ($z = -4.183$, $p < .001$) or /ŋ/ ($z = -3.788$, $p < .001$) whereas accuracy did not differ significantly between /n/ and /ŋ/ ($z = -.400$, $p = .916$).

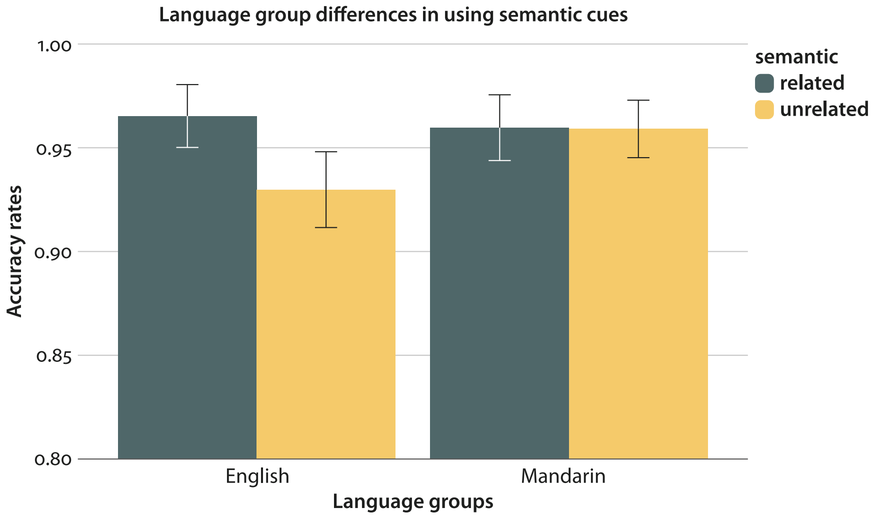


Figure 2. Interaction between language group and semantic relatedness in Experiment 2

Table 5. Results from the linear mixed-effects models predicting log RT and accuracy in Experiment 2

Log RT						
Fixed effects	**Sum Sq**	**Mean Sq**	**Num DF**	**Den DF**	***F*-value**	**Pr(>F)**
Language	.094	.094	1	64.6	31.859	<.001
Coda	.195	.097	2	2022.7	32.734	<.001
Semantic	.094	.094	1	2374.6	31.725	<.001
TO	.012	.006	2	19.8	2.043	.156
Language × TO	.033	.016	2	3635.4	5.489	.004
Accuracy						
Fixed effects	**Sum Sq**	**Mean Sq**	**Num DF**	**Den DF**	***F*-value**	**Pr(>F)**
Language	.041	.041	1	64.6	.918	.341
Coda	.953	.476	2	3432	10.677	<.0001
Semantic	.262	.262	1	3428.9	5.863	.016
Language × Semantic	.212	.212	1	3428	4.747	.029

Note. Language = English vs. Mandarin. Coda = /n/, /ŋ/, vs. /s/. Semantic = semantically related vs. semantically unrelated. TO = trial order.

Similar to Experiment 1, a separate set of models was built to examine the effect of proficiency in nonnative listeners' use of semantic relatedness by including only data from L2 learners with the fixed effects of semantic, coda, proficiency, TO and interaction terms between these variables and random effects of subjects and items. For both accuracy and RT data, proficiency does not interact significantly with either semantic or coda cues (all *p*s > .1).

The results were overall consistent with our hypothesis, that Mandarin L2 learners would not be able to utilize semantic cues to the same extent as native English listeners at the accuracy level. In addition, the semantic effect did not interact significantly with trial order, suggesting that the use of semantic cues in segmentation was robust throughout all three occurrences of the target word. In the L2 learners-only mixed-effects model, proficiency did not have a significant relationship with the semantic effect, suggesting that L2 competence level has minimal influence on the use of semantic cue in the segmentation of nonnative speech. This result was consistent with previous research focusing on speech segmentation (e.g., Tremblay & Spinelli, 2014, White et al., 2010), suggesting that L2 proficiency is less likely to influence the use of semantic cues to the same extent proficiency influences other aspects of L2 processing (e.g., stress processing: Lin, Wang, Idsardi, & Xu, 2014; language switching: Costa & Santesteban, 2004).

Experiment 2 replicated the results of Experiment 1 and found that both language groups experienced more difficulty segmenting words with coda /s/ than /n/ or /ŋ/, suggesting the effect with /s/ is not stimuli-specific. In addition, both English listeners and L2 learners were faster to identify the target word preceded by coda /ŋ/ than /n/. This finding provides evidence for the use of phonotactic legality in the segmentation of both native and nonnative speech. Since /ŋ/ is not a legal onset whereas /n/ is allowed both syllable-initially and finally in both Mandarin and English, it was easier for both language groups to identify the word boundary immediately after /ŋ/.

General discussion

The current study examined the use of lexical and sublexical cues in L2 segmentation. Within the lexical level, results showed that native listeners used both semantic relatedness at the phrasal level and lexical knowledge at the single word level efficiently. In contrast, L2 learners did not use semantic cues to the same extent as the native listeners at the accuracy level. L2 learners were able to use lexicality as a cue to segment L2 speech, although the lexicality effect was not robust throughout all stimuli. Within the sublexical level, both native and nonnative listeners had more difficulty segmenting words with coda /s/ whereas they were faster to identify word boundary following coda /ŋ than /n/. Since we did not observe any significant interaction between lexical cues and coda cues for either group, this suggests that both language groups used both types of cues independently. Our study adds novel contribution to the increasing body of research in L2 speech segmentation, showing that L2 proficiency has an important influence on the use of lexical knowledge, but not on the use of semantic relatedness, in the identification of word boundaries in continuous speech.

In terms of the use of phonotactic constraints in nonnative segmentation, Mandarin L2 learners' showed patterns similar to those of native English listeners. Since there was no significant interaction between coda cues and language group and cloze test scores did not significantly influence the use of coda cues, it appears that nonnative listeners could quickly develop sensitivity to phonotactic probability in the nonnative language irrespective of their L2 competence. Previous research (e.g., Weber & Cutler, 2006; Al-jasser, 2008) has shown similar results with L2 learners of English with Arabic or German. The current study extended these results to English learners with a non-Indo-European L1 background. Furthermore, Al-jasser (2008) found that L2 learners could improve their ability to exploit phonotactic cues in segmentation of L2 speech after eight weeks of training. The participants in the current study had a mean length of U.S. residence of 1.4 years ($SD = 1.37$), suggesting that short amount of immersion or instruction is sufficient for learners to develop sensitivity to the distributional cues in the L2. What is remarkable is that even though /s/ is a not allowed syllable-finally in Mandarin, L2 learners in the current study did not showed additional difficulty identifying a word boundary after /s/. It appears that the cross-linguistic influence of L1 phonological structure in speech segmentation is reduced compared to other areas of L2 speech processing. For example, L2 learners whose L1 does not have lexically contrastive stress had difficulty discriminating pairs of English words that only differ in stress location (Dupoux, Peperkamp, & Sebastián-Gallés, 2001; Lin et al., 2014). Moreover, Japanese L2 learners of English could not discriminate between English r/l due to the fact that Japanese does not differentiate between these two consonants (MacKain, Best, & Strange, 1981). However, with intensive training and long period of immersion in an English-speaking environment Japanese speakers were able to perceive English r/l categorically (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999).

We speculate that Mandarin L2 learners' ability to quickly acquire sensitivity to /s/ as a coda cue may be due to the high likelihood of occurrence of /s/ in the word-final position. For morphologically complex words, -s is an inflectional suffix to indicate third-person singular verbs or plural nouns. In fact, 27% of the 74286 unique words in the SUBTLEXus (Brysbaert & New, 2009) are /s/ final. In contrast, the percentage of unique words in this corpus for /t/, /p/, or /l/-final are 0.8%, 5.8%, and 3.5% respectively. All of these consonants are not allowed in the coda position in Mandarin. It is likely that if the current experiments were replicated with context words ending with /t, p, or l/, Mandarin L2 learners may show a heightened difficulty identifying the word boundary compared to native listeners.

Another possibility is that L2 learners were able to quickly acquire sublexical cues that do not exist in L1 (e.g., coda /s/ do not exist in Mandarin), yet they would apply cues that do exist in L1 to segment L2 speech. For example, Al-jasser (2008) found that Arabic L2 learners of English were faster to spot the target word *line* in the sequence *veebline* in which /bl/ is an illegal onset in Arabic compared to *veefline* in which /fl/ is a possible onset in both languages. If a phonotactic constraint is in conflict between L1 and L2, the nonnative listeners may be more likely to apply L1 phonotactic cues to segment L2 speech. However, if a L2 phonotactic constraint does not exist in the L1, then L2 learners could rapidly develop sensitivity to L2 distributional cues with minimal cross-linguistic influence.

For the use of lexical cues, the current study showed that L2 learners' use of lexical knowledge in segmentation was not as robust as that of native listeners, since the lexical effect was not consistent across all stimuli lists and appearances of the target words. Moreover, learners with higher L2 competence were able to utilize lexical knowledge more efficiently compared to those with lower L2 proficiency. It is likely that learners with higher L2 competence had better vocabulary knowledge in English, therefore, they could rapidly recognize familiar words in the continuous speech (and this recognition may or may not activate the corresponding conceptual representation). Another possibility is that L2 learners with more advanced English proficiency may have less competition from the L1 words. Bilingual word recognition models such as the BIA+ (developed from the Bilingual Interactive Activation Model; Dijkstra & van Heuven, 1998; van Heuven, Dijkstra, & Grainger, 1998) and the Bilingual Language Interaction Network for Comprehension of Speech (BLINCS, Shook & Marian, 2013) assume that the bilingual's two languages share a similar phonological system with a single semantic level with shared conceptual representation; at the lexical level, the two languages are separated but integrated. During speech comprehension, auditory input activates phonological information in both L1 and L2. Although both models assume that L2 proficiency affects lexical activation, studies showed mixed results with regards to the influence of proficiency on phonological access dependent on task design and the degree of phonological overlap in the critical stimuli (e.g., Duyck, Diependaele, Drieghe, & Bryasbaet, 2004; Jared & Kroll, 2001; Lagrou, Hartsuiker & Duyck, 2013; Mishra & Singh, 2016). For example, Zhou et al. (2010) did not find a significant relationship between phonological priming effects and proficiency for Mandarin-English bilinguals involving monosyllabic words from both languages. The context words in the English-only word identification task in the current study were all trisyllabic words, thus, the degree of phonological overlap in the stimuli should be minimal. As a result, L2 learners in the current study may be better at inhibiting the activation of L1 phonology when listening to the auditory context in the current experiments.

Although L2 proficiency has a significant relationship with the use of lexical knowledge in nonnative speech segmentation in Experiment 1, proficiency did not significantly influence the use of semantic relatedness in Experiment 2. It is possible that the cloze test was not sensitive enough to measure participants' understanding of the relationship between word meanings. The cloze test (Bachman, 1982) mainly assesses a language learner's vocabulary knowledge and her ability to comprehend context at the sentence and passage level. Since the semantic relatedness cue measured in Experiment 2 involved semantic relations between two words, the cloze test might not be an accurate measure of the participants' competence in understanding word relations. Future research may consider using a proficiency test that particularly taps into semantic knowledge such as synonym judgment or production tasks used in Perfetti and Zhang (1995) or Niemi, Varuas, and Wright (1980).

One notable limitation of the current study was that a fully counterbalanced design was not employed since each participant saw the same target word three times from two lists. We included sufficient filler items in each experiment to ensure participants did not develop any processing strategy and ensured at least 70 trials separating the appearance of the same target word to minimize the possibility of any repetition priming effect. Nevertheless, future research should increase the number of lists so that each target word only appear once in each list, which would completely eliminate repetition effect.

Conclusion

The present study examined the role of phonotactic and lexical-semantic cues in L2 segmentation by Mandarin learners of English. Although nonnative listeners did not use lexical knowledge and semantic relatedness to the same extent as native listeners, L2 proficiency has a significant relationship with the use of lexical knowledge. In addition, L2 learners showed native-like sensitivity to English phonotactic constraints even though coda /s/ violates the phonotactic legality in Mandarin L1. Our results provided novel evidence demonstrating that L2 learners were able to quickly develop sensitivity to L2 phonotactic probabilities and use coda cues efficiently to segment continuous speech in the nonnative language. Moreover, the current results suggested that L2 learners with advanced proficiency were able to utilize the segmentation strategy of subtraction of known words, despite relatively short length of immersion in the L2-speaking environment. Overall, the cross-linguistic influence of L1 phonological structures on L2 speech segmentation appears to be reduced compared to other areas of L2 processing.

References

Al-jasser, F. (2008). The effect of teaching English phonotactics on the lexical segmentation of English as a foreign language. *System*, 36, 94–106. https://doi.org/10.1016/j.system.2007.12.002

Altenberg, E. P. (2005). The perception of word boundaries in a second language. *Second Language Research*, 21, 325–358. https://doi.org/10.1191/0267658305sr250oa

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese/r/and English/l/and/r. *Journal of Phonetics*, 32(2), 233–250. https://doi.org/10.1016/S0095-4470(03)00036-6

Bachman, L. F. (1982). The trait structure of cloze test scores. *TESOL Quarterly*, 16, 61–7. https://doi.org/10.2307/3586563

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Method*, 39, 445–459. Retrieved from <http://elexicon.wustl.edu/>. https://doi.org/10.3758/BF03193014

Beckman, M., & Edwards, J. (1987). The phonological domains of final lengthening. *Journal of Acoustical Society of America*, 81, S67. https://doi.org/10.1121/1.2024348

Bardovi-Harlig, K. and Dörnyei, Z. (1998), Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning. *TESOL Quarterly*, 32, 233–259. https://doi.org/10.2307/3587583

Bradlow, Ann R., Akahane-Yamada, D. B., & Tokhura, Y. (1999). Training Japanese listeners to identify English/r/and/l: Long-term retention of learning in perception and production. *Attention, Perception, & Psychophysics*, 61(5), 977–985. https://doi.org/10.3758/BF03206911

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

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *Plos ONE*, 5, e10729. https://doi.org/10.1371/journal.pone.0010729

Christie, W. M. (1974). Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America*, 55, 819–821. https://doi.org/10.1121/1.1914606

Church, K. (1983). Allophonic and phonotactic constraints are useful. Presented at International Joint Conference on Artificial Intelligence, Karlsruhe, West Germany.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, 50, 491–511. https://doi.org/10.1016/j.jml.2004.02.002

Cutler, A. (2000). Listening to a second language through the ears of a first. *Interpreting*, 5, 1–23. https://doi.org/10.1075/intp.5.1.02cut

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113–121.

Dilley, L., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language*, 63, 274–294. https://doi.org/10.1016/j.jml.2010.06.003

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness". The Journal of the Acoustical Society of America, 110(3), 1606–1618. https://doi.org/10.1121/1.1380437

Gow Jr., D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 344–359. https://doi.org/10.1037//0096-1523.21.2.344

Hanulikova, A., Mitterer, H., & McQueen, J. M. (2011). Effects of first and second language on segmentation of non-native speech. *Bilingualism: Language and Cognition*, 14, 506–521. https://doi.org/10.1017/S1366728910000428

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, 57, 1–33. https://doi.org/10.1111/j.1467-9922.2007.00397.x

Kroll, J. F. & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149–174. https://doi.org/10.1006/jmla.1994.1008

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition*, 13, 373–381. https://doi.org/10.1017/S136672891000009X

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). *lmerTest: Tests for random and fixed effects for linear mixed effect models* (lmer objects of lme4 package). R package version 2.0-0. <http://CRAN.R-project.org/package=lmerTest>

Lagrou, E., & Hartsuiker, R. J., & Duyck, W. (2013). The influence of sentence context and accented speech on lexical access in second-language auditory word recognition. *Bilingualism: Language and Cognition*, 16, 508–517. https://doi.org/10.1017/S1366728912000508

Lehiste, I. (1960). An acoustic – phonetic study of internal open juncture. *Phonetica*, 5, 5–54. https://doi.org/10.1159/000258062

Lenth, R. V. (2017). Using lsmeans. <https://cran.r-project.org/web/packages/lsmeans/vignettes/using-lsmeans.pdf>

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of speech code. *Psychological Review*, 74, 431–461. https://doi.org/10.1037/h0020279

Lin, C. Y., Wang, M., Idsardi, W. J., & Xu, Y. (2014). Stress processing in Mandarin and Korean second language learners of English. *Bilingualism: Language and Cognition*, 17, 316–346. https://doi.org/10.1017/S1366728913000333

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, 28, 203–208. https://doi.org/10.3758/BF03204766

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research*, 50, 940–967. https://doi.org/10.1044/1092-4388

MacKain, K.S., Best, C.T., & Strange, W. (1981). Categorical perception of English/r/and/l/by Japanese bilinguals. *Applied Psycholinguistics*, 2, 369–390. https://doi.org/10.1017/S0142716400009796
Mattys, S.L. (2004). Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 397–408. https://doi.org/10.1037/0096-1523.30.2.397
Mattys, S.L., White, L., & Melhorn, J.F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134, 477–500. https://doi.org/10.1037/0096-3445.134.4.477
Mattys, S.L., & Melhorn, J.F. (2007). Sentential, lexical, and acoustic effects on the perception of word boundaries. *Journal of the Acoustical Society of America*, 122, 554–567. https://doi.org/10.1121/1.2735105
McClelland, J.L., & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86. https://doi.org/10.1016/0010-0285(86)90015-0
Nakatani, L.H., & Dukes, K.D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, 62, 714–719. https://doi.org/10.1121/1.381583
Newman, R.S., Sawusch, J.R., & Wunnenberg, T. (2011). Cues and cue interactions in segmenting words in fluent speech. *Journal of Memory and Language*, 64, 460–476. https://doi.org/10.1016/j.jml.2010.11.004
Niemi, P., Vauras, M., & Wright, J. (1980). Semantic activation due to synonym, antonym, and rhyme production. *Scandinavian Journal of Psychology*, 21, 103–107. https://doi.org/10.1111/j.1467-9450.1980.tb00347.x
Norris, D.G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52, 189–234. https://doi.org/10.1016/0010-0277(94)90043-4
Norris, D., McQueen, J.M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1209–1228.
Norris, D., McQueen, J.M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, 34, 191–243. https://doi.org/10.1006/cogp.1997.0671
Norris, D., McQueen, J., Cutler, A., Butterfield, S., & Kearns, R. (2001). Language-universal constraints on speech segmentation. *Language and Cognitive Processes*, 16, 637–660. https://doi.org/10.1080/01690960143000119
Perfetti, C.A., & Zhang, S. (1995). Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 24–33. https://doi.org/10.1037/0278-7393.21.1.24
Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research*, 66, 180–194. https://doi.org/10.1007/s00426-002-0086-5
R Development Core Team. (2008). *R: A language and environment for statistical computing*. R Vienna: Foundation for Statistical Computing. <http://www.R-project.org>
Rumelhart, D.E., McClelland, J.L., & PDP Research Group. (1986). *Parallel distributed processing: Exploration in the microstructure of cognition foundations*. Cambridge, MA: The MIT Press.

Sanders, L. D., Neville, H. J., & Woldroff, M. G. (2002). Speech segmentation by native and non-native speakers: The use of lexical, syntactic, and stress-pattern cues. *Journal of Speech, Language, and Hearing Research*, 45, 519–53. https://doi.org/10.1044/1092-4388(2002/041)

Shook, A., & Marian, V. (2013). The bilingual language interaction network for comprehension of speech. *Bilingualism: Language and Cognition*, 16, 304–324. https://doi.org/10.1017/S1366728912000466

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

Umeda, N. (1975). Vowel duration in American English. *Journal of the Acoustical Society of America*, 58, 434–445. https://doi.org/10.1121/1.380688

White, L., Melhorn, J. F., Mattys, S. L. (2010). Segmentation by lexical subtraction in Hungarian speakers of second-language English. *The Quarterly Journal of Experimental Psychology*, 63, 544–554. https://doi.org/10.1080/17470210903006971

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in spoken word recognition. *Psychological Science*, 9, 325–329. https://doi.org/10.1111/1467-9280.00064

Vivevitch, M. S. & Luce, P. A. (1999). Probabilistic phonotacctics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40, 374–408. https://doi.org/10.1006/jmla.1998.2618

Vivevitch, M. S. & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, 36, 481–487. https://doi.org/10.3758/BF03195594

Weber, A. (2000). The role of phonotatics in the segmentation of native and nonnative continuous speech. In A. Cutler, J. M. McQueen, & R. Zondervan (Eds.), *Proceedings of Workshop on Spoken Word Access Processes* (pp. 143–146). Nijmegen: MPI for Psycholinguistics.

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

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America*, 119, 597–607. https://doi.org/10.1121/1.2141003

Zhou, H., Chen, B., Yang, M., & Dunlap, S. (2010). Language nonselective access to phonological representations: Evidence from Chinese – English bilinguals. *The Quarterly Journal of Experimental Psychology*, 63(10), 2051–2066. https://doi.org/10.1080/17470211003718705

Address for correspondence

Candise Yue Lin
University of Southern California
SGM 501, 3620 McClintock Ave.
Los Angeles CA 90089-1061
USA
candisel@usc.edu

Co-author information

Min Wang
University of Maryland
Department of Human Development and Quantitative Methodology
College of Education
minwang@umd.edu