

# The Role of Lexical Knowledge and Stress Cues in Segmentation in Second Language Learners of English

Candise Y. Lin, Min Wang

Department of Human Development and Quantitative Methodology, University of Maryland,  
College Park, U.S.A.

candisec@umd.edu, minwang@umd.edu

## Abstract

This study compared the use of lexical knowledge and stress cues in segmentation by Mandarin second language (L2) of English. Previous research has shown that native English speakers reliably use lexical cues but not stress cues. However, L2 learners may have difficulty using lexical knowledge in segmentation due to their limited vocabulary size. Instead, Mandarin L2 learners may rely on stress cues since Mandarin and English are similar in terms of their stress patterns. Using a cross-modal priming task, results showed that both the native listeners and L2 learners responded faster to initial-stressed target words than final-stressed target words, showing evidence to the use of stress cues in segmentation. No evidence of the use of lexical knowledge was found. It appears that there is cross-linguistic influence of the use of stress cues in segmentation by L2 learners.

**Index Terms:** stress, segmentation, second language learning, spoken word recognition, lexical knowledge

## 1. Introduction

In spoken word recognition, every listener faces the problem of identifying word boundaries from continuous speech stream. Unlike written text where there are visual spaces between words, there is often no pause between words in the spoken input. There are a number of cues listeners can utilize to solve this problem. These cues can be pragmatic, semantic, lexical, acoustic, phonetic, and prosodic. Listeners may have an implicitly ranked order in terms of the importance of each cue depending on language typology and listening condition.

### 1.1. The Hierarchical Framework

Mattys, White, and Melhorn [1] have proposed a hierarchical framework that captures the relative importance of various cues in speech segmentation in English. There are three tiers in this hierarchy (Figure. 1). Tier I represents the lexical cues that encompass lexico-semantic knowledge and sentential context such as pragmatics and syntax. Tier II represents the segmental cues which include phonotactics and acoustic-phonetics such as coarticulation and allophony. Finally, Tier III represents the prosodic cues that include both metrical stress and lexical stress. Mattys et al. [1] suggested that listeners always use the cues from Tier I under the optimal listening condition. When lexical information is poor due to noise in the input, listeners would resort to the segmental cues. Listeners would only use the prosodic cues when both the lexical and segmental cues are not interpretable. In other words, listeners may assign rank-ordered weights to lexical, segmental, and prosodic information when segmenting continuous speech stream to recognize spoken words. This hierarchy was constructed based on findings from native English speakers. Thus, it is unclear whether the same ranked order would apply to segmentation by nonnative speakers of

English. The current study examined segmentation in Mandarin L2 learners of English by competing stress cues against lexical knowledge in a cross-modal priming experiment.

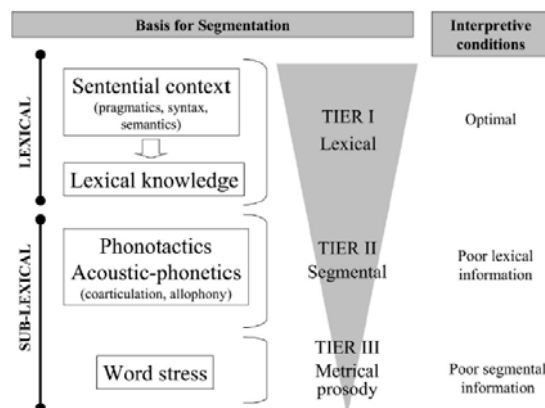


Figure 1. The hierarchical framework proposed by Mattys et al. (2005)

One of the advantages of using lexical knowledge in segmentation is that knowing the boundaries of a recognized word allows listeners to hypothesize the immediately following phoneme to be the beginning of a new word and the immediately preceding phoneme to be the ending of another word [2]. Since there is evidence suggesting that the location of stress is activated post-lexically [3], stress becomes less important if the target word has already been segmented and recognized. However, L2 learners may have to accumulate a relatively large L2 vocabulary size in order to use lexical knowledge efficiently.

In terms of the use of prosodic cues, previous studies have shown that native English speakers often consider a strong syllable (a syllable with an unreduced vowel) as the beginning of a new word and automatically initiate lexical access [4]. For example, the word *today* would result in longer recognition time because listeners may initially group the first syllable *to* with the preceding word and segment *day* from the input. English is a free stress language and the location of stress is unpredictable. In contrast, in fixed stress languages such as Czech and Finnish in which stress is word-initial, stress can be utilized as a reliable cue for segmentation since stress marks the beginning of a word. However, the variability of stress location in English may render stress an unreliable segmentation cue and thus it is ranked the lowest in the hierarchy whereas stress may be ranked higher if the hierarchy was built based on data from Czech or Finnish speakers. Therefore, the ranked order of the cues may be re-arranged depending on language typology.

## 1.2. Segmentation in Second Language Learners

The validity of the hierarchical framework proposed by Mattys et al [1] has only been examined in native Hungarian speakers who are L2 learners of English [2]. Participants were divided into two groups, beginning and intermediate learners, to examine the research questions whether L2 learners' use of lexical cues differs by their level of L2 proficiency. In Hungarian, stress placement is always word-initial. The predictability of stress location in Hungarian may render stress a more reliable segmentation cue in the participants' L1. Therefore, the second research question in this study is whether the presumed heavier weight of stress segmentation in Hungarian would transfer when Hungarian speakers segment speech in L2 English.

White et al. (2010) utilized the cross-modal priming paradigm. In each trial, participants were asked to listen to a five-syllable phrase (e.g., *anythingcorri*) with visual presentation of a three-syllable letter string (e.g., *corridor*) 100ms after the offset of the auditory prime. The participants' task was to determine whether the visual stimulus was a real English word. The first three syllables in the auditory phrase were referred to as the *context* (e.g., *anything*) while the last two syllables were referred to as the *prime* (e.g., *corri*). The critical manipulations were the lexicality of the context (e.g., *anything* or *imoshing*), the stress pattern of the context (e.g., strong-weak (SW) *anything* or (WS) *another*), and the stress pattern of the prime (e.g., *corridor* or *confusion*).

Results showed that both the native English listeners and the Hungarian L2 learners responded faster to target words following real word contexts than nonword contexts. The magnitude of priming did not significantly differ between SW and WS regardless of the lexicality of contexts. These results suggested that both native and nonnative listeners used lexical knowledge in segmentation. The absence of an advantage in SW stress primes also suggested that both groups did not use metrical segmentation as suggested by Cutler and Norris [3]. Furthermore, L2 speakers were divided into four groups based on their performance on the proficiency test. The researchers did not find a lexical priming effect in the lowest proficiency group, suggesting that L2 speakers with a small vocabulary size did not utilize lexical strategy in segmentation.

White et al. [2] concluded that the hierarchical framework can also be generalized to Hungarian L2 learners who were able to exploit lexical cues in segmentation. The absence of any stress effect is also consistent with the hierarchy which predicts that listeners only resort to the prosodic cues in degraded listening conditions. However, Hungarian learners may not use stress cues in segmentation not because they are ranked lower than lexical cues but because Hungarian listeners do not encode stress in their phonological representation. Previous research has shown that speakers of fixed stress languages, such as Finnish and Turkish, cannot discriminate minimal stress pairs if the task prevents them from using acoustic cues (e.g., *pitch*, *intensity*, and *duration*) by imposing a high demand on working memory [5]. Thus, it is possible that Hungarian learners of English could not utilize stress segmentation because they do not have an abstract representation of stress in phonological memory.

The findings reported by White et al [2] are also inconsistent with those from Endress and Hauser [6]. Using a word learning paradigm, Endress and Hauser found that monolingual English listeners were able to segment Hungarian words they learned from Hungarian sentences they had never heard before. Although there is no vowel reduction in

unstressed vowels in Hungarian, both Hungarian and English use pitch, duration, and intensity to realize stress. Endress and Hauser suggested that prosody may be a universal segmentation cue. Endress and Hauser's claim is not inconsistent with the explanation of stress representation. English listeners may be able to use stress cues to segment a foreign language because they encode stress in memory [6]. However, since representation of stress is not available to Hungarian learners of English, they could not utilize stress cues to segment an L2 that has minimal stress pairs. Thus, stress cues may only be universal in segmentation if the listener has a phonological representation of stress.

## 1.3. Segmentation in Mandarin Learners of English

The stress system in Mandarin is similar to that in English. Stress is lexically contrastive in both languages and it has been shown empirically that native speakers of these languages encode stress in phonological memory [7]. The realization of stress involves a similar set of acoustic correlates including pitch, duration, intensity, and change in vowel quality [8, 9]. However, stress location in Mandarin is relatively more predictable than that in English. A weak syllable can never be word-initial in Mandarin. In contrast, an unstressed syllable can take any position in an English word (e.g., *support* /sə'pɔrt/, *catalog* /'kætə,lɒg/, *museum* /mju:'ziəm/). Thus, a stressed syllable is more likely to coincide with the beginning of a word in Mandarin. As a result, stress may be a more reliable segmentation cue in Mandarin and ranked higher in Mandarin listeners' implicit hierarchy. We hypothesized that Mandarin L2 learners of English would show cross-linguistic influence in their use of segmentation cues so that Mandarin listeners would segment initial-stressed words faster than medial-stressed words from context. We also hypothesized that, based on Mattys et al.'s findings [1], native English listeners would not show a stress effect. In terms of the use of lexical cues, since the participants were current university students in the U.S. and immersed in an English-speaking environment, we hypothesized that the L2 learners, as well as the native English listeners, would show faster response latency to target words following real word contexts than those following nonword contexts.

## 2. Method

### 2.1. Participants

All participants were current students at a mid-Atlantic university in the U.S. As of today, 10 Mandarin L2 learners of English and 10 native English listeners have been tested. Data collection will continue to ensure larger sample size in each language group. To assess the participants' English proficiency, they completed the Language Experience and Proficiency Questionnaire (LEAP-Q) [10] and a cloze test (a standardized measurement of English proficiency in the form of a fill-in-the-blank activity) (Table 1).

Table 1. Demographic profiles for the language groups. P-values obtained by t-tests comparing group means.

Demographic	English	Mandarin	p-value
Mean Age (years)	24.6	24.2	.781
Self-rated listening proficiency (1 lowest and 10 highest)	9.8	7	.000
Cloze (accuracy)	91.4%	75.4%	.003

## 2.2. Measures

We adapted the cross-modal priming lexical decision task from Mattys et al. [1]. Ninety-six pairs of initial-stressed and medial-stressed trisyllabic monomorphemic words were selected from a list generated from the English Lexicon Project (retrieved from <http://elexicon.wustl.edu/>). Each word in the pair was closely matched on surface frequency based on the norms from the Hyperspace Analogue to Language (HAL) database. This list was sent to ten Mandarin L2 learners of English in the same population where the current sample was drawn from. They were asked to rate how familiar they are 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 5 were selected to ensure that the L2 learners know the words. The final stimuli list consisted of 20 pairs of words matched on the rhyme of the final syllable that served as the contexts and 20 pairs of words matched on the onset of the first syllable that served as the primes. Each word in the pair was matched on written and spoken frequency, familiarity, number of letters, number of phonemes, biphone frequency, size of phonological neighborhood and uniqueness point.

Nonword contexts were created using the Phonotactic Probability calculator (website can be accessed from <http://www.people.ku.edu/~mvitev/PhonoProbHome.html>) so that each phoneme in the nonword is matched with the phoneme in the corresponding real word in terms of position-specific probability. Twenty pairs of nonwords were created and each pair consisted of an initial-stressed word and a medial-stressed word.

The design was a 2 x 2 x 2 factorial design with the three factors being the lexicality of the context, the stress pattern of the context, and the stress pattern of the prime (Table 2). This yielded eight possible combinations of the test phrase for each of the 20 pairs of contexts and primes. A test phrase was made up by the context and the first two syllables of the prime word. To prevent repeated exposure to the same prime or context, we selected four test phrases from each set of eight and the selection was counterbalanced on the three factors. Two lists were created so that List 1 contains four test phrases from one set and List 2 contains the other four test phrases from the same set. All participants heard 80 experimental utterances.

Table 2. Examples of the critical stimuli.

Context				Prime	
Word		Nonword		SW	WS
SW	WS	SW	WS		
character	consider	manister	dilieter	register	remember
battery	attorney	bulnerty	umony	recipe	republic
cinema	umbrella	tarila	elboma	politic	pajama

Priming effects were determined by comparing the reaction time to the test phrase and that to a baseline phrase. The baseline phrase was created to match each of the test phrases by replacing the prime (e.g., the last two syllables of the test phrase) with a distorted speech of matched duration. The distorted speech was created by filtering the speech through a noise band vocoder with one channel using Tigerspeech (a speech processor that simulates the hearing condition of cochlear-implant users). Two different sets of 80 baseline phrases were created for List 1 and List 2.

The same female native English speaker recorded all stimuli. She pronounced each full phrase without interruption (e.g., *considerremember*). After recording, the last syllable of the phrase was manually cut out, leaving the five-syllable test phrase (e.g., *considerremem*). In each experimental trial, the visual prime (e.g., *remember*) was presented 100ms after the offset of the auditory test phrase (e.g., *considerremem*) or the baseline phrase. To prevent participants from developing processing strategies, three types of fillers were created. The first type of fillers consisted of all nonword visual targets to balance the number of “yes” and “no” responses. Similar to the experimental trials, the nonword targets were presented 100ms after the offset of the auditory phrase. The second type of fillers consisted of half nonword and half real word targets which were presented immediately after the offset of the third syllable of the auditory phrase. The third type of fillers consisted of half real word and half nonword targets which were presented immediately after the offset of the second syllable of the auditory phrase. There were 160 trials in the first type of fillers and 100 trials in each of the second and third types of fillers. All filler trials were equally divided between those with real word or nonword contexts and those with congruent or incongruent (e.g., distorted speech) primes.

Each experiment consisted of a total of 520 trials, equally divided into four blocks of 130 trials each. The presentation of trials in each block is pseudorandomized so that there are at least 40 trials between two repeated primes or contexts. We also ensured that there were no more than three consecutive real word or nonword targets in a row. The order of blocks was counterbalanced across participants.

## 2.3. Procedure

Participants were tested individually using a desktop PC in a quiet room. They were randomly assigned to List 1 or List 2. Each participant completed all four blocks with a 5-minute break between each block. The experiment was implemented via the E-prime software. For each trial, participants heard the auditory phrase via headphones and saw a sequence of letter strings visually presented on the center of the screen written in 22pt bold Courier font. They were instructed to decide whether the letter strings constitute a real English word or not by pressing the keys labeled “Yes” or “No” on the computer keyboard. Speed and accuracy was emphasized. Participants received 10 practice trials before the actual experiment.

## 3. Results

Only data from the 80 critical trials and 80 baseline trials were analyzed. Response time (RT) data for incorrect responses and data that are two standard deviations above or below the cell mean were deleted from analysis. This results in 9.9% deletion of the original data. Both native and nonnative listeners had accuracy higher than 85% across all conditions. RT data were analyzed using repeated-measure ANOVA with Condition (baseline vs. priming), Context Lexicality (word vs. nonword), Stress of Context (SW vs. WS) and Stress of Prime as within-subject factors and Language Group as a between-subject factor. Preliminary results from the RT data showed a significant priming effect for Condition ( $F(1, 18) = 18.707$ ,  $p < .000$ ). All participants responded faster to target words following the prime than following the baseline. There is a significant main effect of Language Group ( $F(1, 18) = 13.337$ ,  $p = .002$ ). The Native English listeners responded significantly faster than the Mandarin L2 learners. There is a significant

interaction between Condition and Stress of Prime ( $F(1, 18) = 4.717, p = .043$ ) (Figure 2). When the prime is congruent with the target, RT is significantly faster when the prime is a medial-stressed word than when the prime is an initial-stressed word ( $t(19) = 2.631, p = .016$ ). When the prime is incongruent with the target (e.g., baseline), there is no significant difference in RT regardless of the stress location of the prime ( $t(19) = -1.221, p = .237$ ).

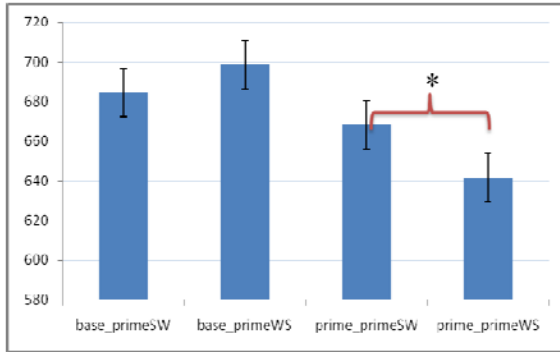


Figure 2. Interaction between Condition and Stress of Prime in Response Time,  $*p < .05$

There is also a significant interaction between Language Group and Stress of Prime ( $F(1, 18) = 4.997, p = .038$ ) (Figure 3). Planned pairwise comparison revealed that in the Mandarin group, there is no significant difference in RT regardless of the stress location of the prime ( $t(9) = -.861, p = .412$ ). However, the native English listeners responded significantly faster to medial-stressed primes than initial-stressed primes ( $t(9) = 2.329, p = .045$ ). There is a marginally significant interaction between Context Lexicality and Stress of Prime ( $F(1, 18) = 3.584, p = .075$ ). Post hoc analyses showed that when the context is a real word, both native and nonnative listeners responded significantly faster to medial-stressed primes than to initial-stressed primes ( $t(19) = 2.297, p = .033$ ). When the context is a nonword, RT did not differ regardless of the stress location of the prime. All other main effects and interactions are not significant (all  $ps > .1$ ).

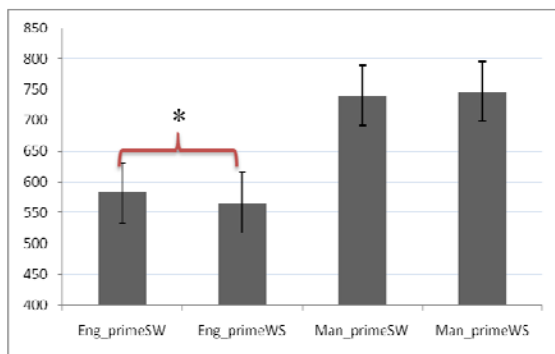


Figure 3. Interaction between Language Group and Stress of Prime in Response Time,  $*p < .05$

#### 4. Discussion

The current study compared the use of stress cues and lexical knowledge in segmentation by native English listeners and

Mandarin L2 learners. Our preliminary results showed that both native and nonnative listeners made faster lexical decision for medial-stressed words than for initial-stressed words when the prime is congruent with the target or when the context is a real word. These results implied that native and L2 listeners were able to utilize lexical and stress cues in segmentation. It is surprising that English listeners responded faster to medial-stressed words even though medial-stress does not coincide with word boundaries. However, in the cross-modal priming task, medial-stress does coincide with the right edge of the auditory phrase because the last syllable was cut off from the prime (e.g., the last syllable in *remember* was manually deleted to form *considerremem*). Thus, medial-stressed words facilitate lexical access as a result of the design of the experiment. Although stress location is relatively less predictable in English than in Mandarin, both groups showed evidence of utilizing stress cues in segmentation and this result is not consistent with White et al.'s [2] study with Hungarian L2 learners of English. Perhaps listener must be able to encode stress in phonological representation in order to use stress cues. The marginally significant interaction between the lexicality of context and the stress pattern of the prime provided tentative evidence for the integration of multiple cues during segmentation and lexical access. This is not consistent with the hierarchical framework because if lexical knowledge is indeed ranked higher than stress cues in segmentation, we should not have observed a significant stress effect when lexical cues are available in the speech signal. If listeners can identify word boundary based on lexical cues, there is no need to utilize stress cues which is at the bottom of the hierarchy. Considering our small sample, we must be cautious when interpreting these results. It is evident that stress cues are used by both native and nonnative listeners but it is still inconclusive whether lexical knowledge is ranked higher than stress cues. However, the fact that we could obtain significant findings with a limited sample size is encouraging. We are hopeful that with an increased sample size, we will be able to obtain more reliable results.

#### 5. Acknowledgements

The current study was supported by the NSF IGERT award DGE-9791455 to the University of Maryland.

#### 6. References

- [1] Mattys, S. L., White, L., & Melhorn, J. F., "Integration of multiple speech segmentation cues: A hierarchical framework". *Journal of Experimental Psychology: General*, 134: 477-500, 2005.
- [2] White, L., Melhorn, J. F., & Mattys, S. L., "Segmentation by lexical subtraction in Hungarian speakers of second-language English". *The Quarterly Journal of Experimental Psychology*, 63: 544-554, 2010.
- [3] Cutler, A., & Norris, D., "The role of strong syllables in segmentation for lexical access". *Journal of Experimental Psychology: Human Perception and Performance*, 14: 113-121, 1988.
- [4] Gutierrez-Palma, N. & Palma-Reyes, A., "On the use of lexical stress in reading Spanish". *Reading and Writing*, 21: 645-660, 2008.
- [5] Dupoux, E., Peperkamp, S., & Sebastian-Galles, N., "A robust method to study stress 'deafness'". *Journal of Acoustical Society of America*, 110: 1606-1618, 2001.
- [6] Endress, A. D., & Hauser, M. D., "Word segmentation with universal prosodic cues". *Cognitive Psychology*, 61: 177-199, 2010.
- [7] Lin, C. Y., Wang, M., Idsardi, W. J., & Xu, Y., "Stress perception in Mandarin-English and Korean-English bilinguals". Under review, 2012.
- [8] Fry, D. B., "Duration and intensity as physical correlates of linguistic stress". *JASA*, 27: 756-768, 1955.
- [9] Duanmu, S., "The phonology of standard Chinese, second edition", Oxford University Press, 2007.
- [10] Marian, V., Blumenfeld, H. K., & Kaushanskaya, M., "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, 2007.