L2 SPEECH SEGMENTATION FOR WORD RECOGNITION: THE ROLE OF LEXICAL STRESS AND SYLLABLE STRUCTURE

Juan J. Garrido-Pozú

Furman University
juan.garrido@furman.edu

ABSTRACT

Current understanding of speech perception suggests that the routines that listeners use to segment the speech signal for word recognition are language specific. English speakers exploit stress patterns for lexical segmentation whereas Spanish speakers use syllable structure. This study investigated L2 lexical segmentation and specifically examined whether L2 learners with a syllable-based L1 (Spanish) exhibit stress-based segmentation in L2 English. 81 English–Spanish learners and 72 Spanish–English learners completed a word-spotting task that manipulated lexical stress and syllable structure. Participants aurally identified English monosyllabic words embedded in disyllabic pseudowords.

Results showed that Spanish-English learners transferred L1 segmentation routines to L2 English. Spanish-English learners exhibited syllable-based segmentation of English regardless of differences in lexical stress, which was more prominent as L2 proficiency increased. English-Spanish learners showed stress-based segmentation regardless of differences in syllable structure. This study suggests that listeners develop only the segmentation routines that correspond to the phonological composition of their L1.

Keywords: L2 speech segmentation, L2 speech perception, lexical stress, syllable.

1. INTRODUCTION

For successful communication, listeners need to segment the speech stream to identify where words begin and end. Word boundaries do not emerge implicitly in the speech signal, but they are rather imposed by the listener using explicit procedures [2]. Each language has specific features that can be exploited for speech segmentation. Early studies found that statistical correlates of the sound patterns of words can be useful for speech segmentation due to the co-ocurrence of syllables within a word being more frequent than those across words [3, 12, 13]. In addition, phonetic cues specific to pre and post boundary phonemes can be exploited for segmentation [4, 17, 20]. In English, for example, voiceless stops are aspirated at word onsets but unaspirated at word offsets. Such cues can be useful for word boundary identification. Listeners can also use implicit knowledge of the structural restrictions of syllables in a particular language to identify the sequences that are permissible as onsets and offsets. Using syllable patterns is notably more efficient for segmentation of languages that have more consistent syllable structure, like Spanish, but not languages that have more variable syllable patterns, like English. In English, however, suprasegmental information provides reliable cues for segmentation [10].

2. STRESS-BASED SEGMENTATION

Prosodic features such as stress and rhythm can serve as cues for the identification of word boundaries in a language like English. The Metrical Segmentation Strategy (MSS) [5, 10] claims that placing word boundaries before stressed syllables is an efficient segmentation strategy that works for at least 90% of content words in English because most content words begin with a metrically stressed syllable. This strategy works for content words that begin with a stressed syllable, but not for content words that begin with an unstressed syllable or for function words. Consequently, the MSS assumes two separate strategies: content words are detected through stressed initial syllables whereas function words are detected through unstressed initial syllables [6].

Statistical and distributional analyses of the structure of the vocabulary of English supports the efficacy of treating strong syllables as cues for word onsets. Cutler & Carter [6] reported that in a computer-readable English dictionary with over 33,000 word entries, 73% of the words had a strong initial syllable and only 27% had a weak initial syllable. Cutler & Carter [6] analyzed a corpus of natural speech samples containing over 190,000 words from British English conversations. Over 90% of these words had a strong initial syllable. Clearly, words with strong initial syllables are more frequent than words with weak initial syllables in average speech contexts, which supports the reliability of strong syllables as cues for word onsets [5].

More evidence in favor of the MSS is found in Cutler & Norris [10]. English monolinguals completed a word-spotting task in which they listened to real monosyllabic words (e.g., *mint*) embedded in disyllabic nonwords that had either two strong syllables (e.g., *mintayf*) or one strong and one weak syllable (e.g., *mintef*). Results revealed that listeners detected real words faster when the second syllable of the nonword was weak than when it was strong. Listeners had difficulty detecting *mint* in *mintayf* because they placed a boundary before the second strong syllable (-*tayf*) and treated it as the beginning of a new word. In *mintef*, on the other hand, there was no segmentation within the word because the second syllable is weak, facilitating recognition of the target *mint*.

3. L2 STRESS-BASED SEGMENTATION

The type of segmentation strategies that listeners use for a L2 may be constrained by the nature of their L1. Studies have reported that bilinguals with a syllable-based L1 like Italian or Spanish can develop L2 segmentation routines that are not syllable-based. However, bilinguals with a stress-based L1 like English struggle to develop syllable-based routines for a L2. In Cutler et al. [9], French-English and English-French early bilinguals completed two syllable monitoring tasks (in English and French) and one word-spotting task in English. Only the French-dominant bilinguals exhibited syllable-based segmentation in the French syllable-monitoring task, and no syllabic effects were observed in the English syllable-monitoring task. In the word-spotting task, only the English-dominant group showed stress-based segmentation. Bilinguals exhibited only the segmentation routine that was motivated by the phonology of their dominant language. French-dominant participants behaved like French monolinguals, and English-dominant participants behaved like English monolinguals; however, the lack of syllabic effects for French-dominant bilinguals with English words shows that these bilinguals may have employed two different segmentation routines, a syllabic one for their dominant language (French) and a non-syllabic one for their non-dominant language (English).

Studies with L2 learners of English with different L1s have reported that L2 learners exhibit evidence of stress-based segmentation. In Katayama [15], the results of a syllable monitoring task revealed that English monolinguals and Japanese learners of English identified target syllables faster in disyllabic words with initial stress than with final stress, which is evidence of stress-based segmentation. Similarly, in Cutler & Shanley [11], Arabian and Chinese learners of English exhibited stress-based segmentation in a word-spotting task in English where they detected monosyllabic words more accurately when they were embedded into disyllabic pseudowords with initial stress than with final stress. These studies showed that L2 learners who speak a L1 that is not stress-based still exhibited L2-specific stress-based segmentation, suggesting that L2 learners can develop segmentation strategies that are not necessarily motivated by L1 phonology. However, it is still unclear whether L2 learners of English with a syllable-based L1 like Spanish can develop L2 stress-based segmentation. Notably, how segmentation routines change as L2 proficiency increases also needs further investigation.

4. THE PRESENT STUDY

The present study investigated whether L2 learners with a syllable-based L1 (Spanish) exhibit evidence of L2 stress-based segmentation. The specific goals of the present study are to test whether L2 learners exhibit stress-based segmentation of English, to test whether L2 learners exhibit syllable-based segmentation of English, and to assess whether L2 segmentation changes as L2 proficiency increases. The study is driven by the following research questions:

1. Are L2 learners sensitive to lexical stress during segmentation of English?
2. Are L2 learners sensitive to syllable structure during segmentation of English?
3. Does L2 proficiency modulate learners' sensitivity to lexical stress and/or syllable structure during segmentation of English?

5. METHODOLOGY

5.1. Participants

The participants included 81 English L1 learners of Spanish and 72 Spanish L1 learners of English. The English L1 subjects were born and raised in the United States and had no intensive exposure to L2 Spanish before puberty. The Spanish L1 subjects were born and raised in Peru and had no intensive exposure to L2 English before puberty. The mean age of English L1 subjects was 22.2 (*SD* = 5.06) and of Spanish L1 subjects was 31.4 (*SD* = 6.76). English L1 subjects started L2 instruction at age 11.9 (*SD* = 6.33), studied L2 Spanish for 8.51 years (*SD* = 3.82), and were exposed to L2 Spanish 18% of the time per week (*SD* = 15.3). Spanish L1 subjects started L2 instruction at age 15.2 (*SD* = 5.12), studied L2 English for 4.67 years (*SD* = 4.03), and were exposed to L2 English about 31.6% of the time per week (*SD* = 20.7). All participants completed the LexTALE [18] and the LexTALE-ESP [14] which provided a measure of participants' proficiency in English and Spanish respectively.

5.2. Stimuli and procedure

Participants completed four tasks in the following order: a background questionnaire, a word-spotting task, the LexTALE test, and the LexTALE-ESP test. Participants completed all tasks in one session which lasted about 40 minutes.

5.2.1. Word-spotting task

The word-spotting task (WST) followed the design in Cutler et al. [9]. Participants listened to isolated English pseudowords (e.g., *mintef*, *mintayf*) and indicated whether the pseudowords contained a real English word embedded at the beginning (e.g., *mint* in *mintef*, *jump* in *jumpev*). In each trial, participants first saw the question "*Does it begin with a real word?"* at the top center of the screen and the words *YES* and *NO* on each side of the screen. 500 ms later, participants heard a pseudoword (e.g., *mintef*). They pressed the left arrow key to answer *YES* and the right arrow key to answer *NO*. If participants answered *YES*, they were asked to type the real English word that they had identified. The task recorded accuracy and response times.

Participants listened to a total of 104 disyllabic pseudowords: 64 experimental items, and 40 filler/distractors. The 64 experimental items began with a monosyllabic real English word. Half of the experimental items contained a real word with a CVCC structure (e.g., *mint*, *risk*) while the other half had a CVC structure (e.g., *thin*, *kiss*). Real words were made into pseudowords by adding segments to the end. CVCC words were attached to a final VC sequence with either a strong or weak syllable (e.g., *mint*: *mintayf*, *mintef*; *risk*: *riskeeb*, *riskeb*). CVC words were attached to a final CVC sequence with either a strong or weak syllable (e.g., *thin*: *thintayf*, *thintef*; *kiss*: *kisskeeb*, *kiskeb*). CVCC pseudowords were matched with a CVC pseudoword with which they shared the same ending. Fillers and distractors consisted of disyllabic pseudowords without any embedded real words. The pseudowords were recorded by a female native speaker of US English from New Jersey. The speaker was instructed to read the items at a normal rate and did not have knowledge of the target words. Target words were not recorded as list initial or list final to avoid intonational differences.

5.3. Data analysis

The data were analyzed using mixed effects models. The WST measured response times (RTs) and accuracy. RTs were analyzed using linear mixed effects models while correct responses were analyzed using logistic mixed effects models. The independent variables were coded as *stress type* (initial, final), *word type* (CVC, CVCC), and *proficiency* (treated as a continuous variable). The models had random intercepts by participant and by item and included all interactions. Main effects and interactions were assessed by partitioning the variance hierarchically via nested model comparisons. The statistical analyses were carried out using R [22]. The analyses used the packages *lme4* [1], *lmerTest* [16], and *emmeans* [19].

6. RESULTS

For English L1 subjects, there was a main effect of *stress type* on the proportion of correct responses ($\chi2(1) = 3.87$, $p = .04$). English L1 subjects identified more accurately words embedded in pseudowords with initial stress than final stress (see Fig. 1). The log odds of identifying the target word correctly when the pseudoword had final stress were 1.07 (β: 1.07, $SE = 0.22$, $p < .001$), but when the pseudoword had initial stress, the log odds increased by 0.51 (β: 0.51, $SE = 0.26$, $p = .04$). The models also revealed a main effect of *stress type* on RTs ($\chi2(1) = 4.49$, $p = .03$). Overall, English L1 subjects identified real words embedded in pseudowords with initial stress about 169ms faster than those in pseudowords with final stress ($SE = \pm.08$) (see Fig. 2). There was not an effect of *word type* on the proportion of correct responses or RTs of English L1 subjects.

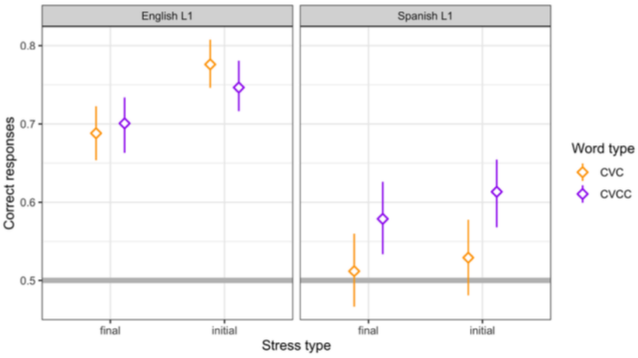

Figure 1: Proportion of correct responses of English L1 subjects (left panel) and Spanish L1 subjects (right panel) as a function of stress type and word type.

For Spanish L1 subjects, there was a main effect of *proficiency* ($\chi2(1) = 21.03$, $p < .001$) and *word type* ($\chi2(1) = 4.06$, $p = .04$) on the proportion of correct responses. There was also a two-way interaction between *word type* and *proficiency* ($\chi2(4) = 4.06$, $p < .04$). Spanish L1 subjects identified more accurately real words with a CVCC structure than with a CVC structure (β: 0.45, $SE = 0.22$, $p = .04$) (see Fig. 1 and 3). Overall, the log odds of identifying the target word correctly increased by 0.8 as English L2 proficiency increased ($SE = \pm.16$). Syllable structure played a role at higher levels of L2 proficiency but not much at lower levels (see Fig. 3). The models yielded no effect of *stress type* on the proportion of correct responses of Spanish L1 subjects and no main effects or interactions on RTs. On average, the Spanish L1 group exhibited slower responses than the English L1 group overall (see Fig. 2), a difference of approximately 830 milliseconds (*SE* = ±.115) and lower rates of correct responses (*M* = .56, *SD* = .49) than the English L1 group (*M* = .73, *SD* = .44) (β: -0.93, *SE* = 0.22, *p* < .001) (see Fig. 1).

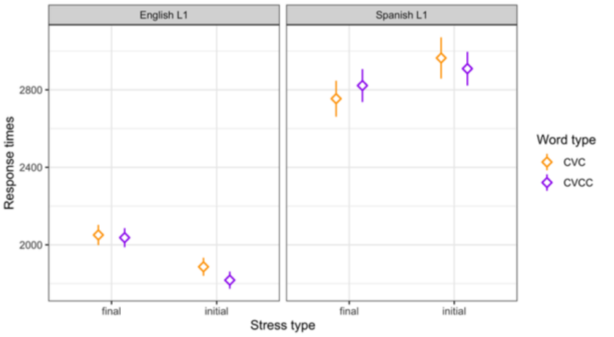

Figure 2: Response times in milliseconds of English L1 subjects (left panel) and Spanish L1 subjects (right panel) as a function of stress and word type.

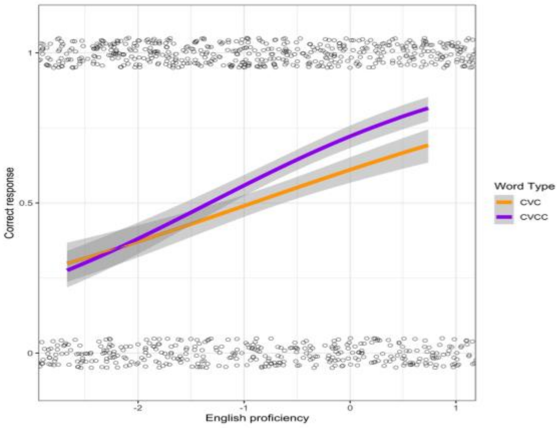

Figure 3: Proportion of correct responses of Spanish L1 subjects as a function of L2 English proficiency and word type.

7. DISCUSSION

This study investigated whether L2 learners develop separate segmentation routines that are specific to the phonology of their L2. Specifically, the study tested whether L2 learners with a syllable-based L1 exhibit L2 stress-based segmentation in a word-spotting task that manipulated lexical stress and syllable structure. The first research question asked whether L2 learners use lexical stress for segmentation of English. Lexical stress affected the responses of English L1 learners of Spanish but not of Spanish L1 learners of English. English L1 subjects identified real words embedded in pseudowords with initial stress faster and more accurately than in pseudowords with final stress, which is considered evidence of stress-based segmentation and coincides with previous studies reporting similar results in English monolinguals [5, 6, 21] and English-dominant bilinguals [8, 9]. The results support the Metrical Segmentation Strategy [10] which suggests that strong syllables can serve as reliable cues for locating possible word onsets. However, the results of this study differ from other studies showing evidence of stress-based segmentation in L2 learners with Japanese L1 [15], Mandarin Chinese L1, and Arabic L1 [11]. Such difference may be due to typological differences between Spanish (syllable-based) and the other languages in question.

The second research question asked whether L2 learners use syllable structure for segmentation of English. Only Spanish L1 subjects showed evidence of syllable-based segmentation of English. The results suggest that learners transfer L1 segmentation routines to a L2, which supports previous studies claiming that L2 segmentation is constrained by L1 segmentation [7, 8, 9]. However, the direction of the syllabic effects was not given as expected. Spanish L1 subjects identified CVCC words more accurately than CVC words, which may indicate that they placed the syllable boundary after the medial consonant cluster (e.g., *thint.ayf* /*thint.ef*; *mint.ayf*/*mint.ef*), facilitating recognition of CVCC words. This syllabification pattern differs from the pattern expected for Spanish and could be evidence of L2-specific syllabification.

The last question focused on the role of L2 proficiency. English L1 subjects did not exhibit syllabic effects even at higher levels of L2 Spanish proficiency. Spanish L1 subjects exhibited stronger syllabic effects at higher levels of English proficiency. Considering the complexity of a word-spotting task, the lack of syllabic effects in lower proficiency learners of English may correspond to difficulties in lexical recognition rather than the absence of a syllable-based segmentation routine.

8. CONCLUSION

This study showed that whereas English L1 learners of L2 Spanish exhibited stress-based segmentation of English, Spanish L1 learners of L2 English displayed syllable-based segmentation of English. The study brings support to the Metrical Segmentation Strategy [10] and suggests that segmentation routines are language specific and L2 segmentation is constrained by L1 segmentation even at higher levels of L2 proficiency.

9. REFERENCES

[1] Bates, D., Mächler, M., Bolker, B., & Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

[2] Carroll, S. E. 2004. Segmentation: learning how to hear words in the L2 speech stream. *Transactions of the Philological Society*, *102*(2), 227–254.

[3] Chomsky, N. 1955. *The logical structure of linguistic theory*. Cambridge, MA: Cambridge University Press.

[4] Christie, W. M. 1974. Some cues for syllable juncture perception in English. *J. Acoust. Soc. Am.*, *55*(4), 3.

[5] Cutler, A., & Butterfield, S. 1992. Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*(2), 218–236.

[6] Cutler, A., & Carter, D. M. 1987. The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, *2*(3), 133–142.

[7] Cutler, A., Mehler, J., Norris, D., & Segui, J. 1986. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, *25*(4), 385–400.

[8] Cutler, A., Mehler, J., Norris, D., & Segui, J. 1989. Limits on bilingualism. *Nature*, *340*(6230), 229–230.

[9] Cutler, A., Mehler, J., Norris, D., & Segui, J. 1992. The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, *24*(3), 381–410.

[10] Cutler, A., & Norris, D. 1988. The Role of Strong Syllables in Segmentation for Lexical Access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 113–121.

[11] Cutler, A., & Shanley, J. 2010. Validation of a training method for L2 continuous speech segmentation. *Interspeech*, 1844–1847.

[12] Harris, Z. S. 1955. From phoneme to morpheme. *Language*, *31*, 190–222.

[13] Hayes, J., & Clark, H. 1970. Experiments in the segmentation of an artificial speech analog. In J. Hayes (Ed.), *Cognition and the development of language* (pp. 221–234). New York: Wiley.

[14] Izura, C., Cuetos, F., & Brysbaert, M. 2014. Lextale-ESP: A test to rapidly and efficiently assess the Spanish vocabulary size. *PSICOLOGICA*, *35*(1), 49–66.

[15] Katayama, T. 2015. Effect of Phonotactic Constraints on Second Language Speech Processing. *I-Perception*, *6* (6).

[16] Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

[17] Lehiste, I. (1960). An acoustic–phonetic study of internal open juncture. *Phonetica*, *5*(s1), 5–54.

[18] Lemhöfer, K., & Broersma, M. 2012. Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343.

[19] Lenth, R. V. 2022. *Emmeans: Estimated marginal means, aka least-squares means*.

[20] Nakatani, L., & Dukes, K. 1977. Locus of segmental cues for word juncture. *The Journal of the Acoustical Society of America*, *62*(3), 714–719.

[21] Norris, D., & Cutler, A. 1988. The relative accessibility of phonemes and syllables. *Perception & Psychophysics*, *43* (6), 541–550.

[22] R Core Team. 2020. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.