# A cross-language investigation of word segmentation by bilinguals with varying degrees of proficiency: Preliminary results.

*Annie C. Gilbert*[1,2,3], *Inbal Itzhak*[1,2,3],*Shari Baum*[1,2,3.]

[1] Center for Research on Brain, Language, and Music, Montréal, Canada.
[2] Montreal Bilingual Brain Initiative, McGill University, Montréal, Canada.
[3] School of Communication Sciences and Disorders, McGill University, Montréal, Canada.

{annie.c.gilbert, inbal.itzhak, shari.baum} @mcgill.ca, www.crblm.ca

## Abstract

An extensive body of research on word segmentation across languages has shown that different languages rely on different cues and strategies to segment meaningful units from the speech stream. These cross-language differences make segmentation difficult for L2 learners, and some previous work showed that bilingual speakers tend to keep applying their L1 segmentation cues to the L2. But bilingual experience varies a great deal, even within a bilingual community, so one might ask if such a pattern applies across all bilinguals regardless of language proficiency, dominance, or everyday use. To investigate this, we designed a cross-modal priming task in which a wide range of English-French bilinguals listened to English and French sentences with ambiguous syllable strings containing either two monosyllabic words (e.g. key we) or one bisyllabic word (e.g. kiwi), produced with context-specific natural prosody. A picture prompt representing either the first monosyllabic word (e.g. a key), or the bisyllabic word (e.g. a kiwi) was presented at the offset of the first syllable of the ambiguous region. Each sentence was presented paired with each picture. Preliminary analyses of a subgroup of English-dominant participants show that they process French and English ambiguous strings differently, and that their segmentation schemes seem to vary with L2 proficiency.

**Index Terms**: bilingualism, word segmentation, individual differences, English, French.

## 1. Introduction

An extensive body of research on word segmentation has shown that different languages rely on different cues and strategies to segment meaningful units from the speech stream. For instance, many studies have shown that speakers of stress-timed languages (like English or Spanish) are sensitive to boundaries between stress units and use lexical stress to locate word onsets in the speech stream [1-5]. On the other hand, speakers of syllable-timed languages like French are sensitive to boundaries between syllables and rely instead on phrase-final lengthening to locate word offsets [1, 6-9]. These cross-language differences can make it difficult for L2 learners, who cannot simply transfer their L1 strategies onto the new language, but instead must learn a new set of cues and adopt a different approach to segmentation. Moreover, the transition from an L1 to an L2 can be more or less difficult depending on the languages' specific prosodic systems. For instance, transitioning from one syllable-timed language to another might be fairly easy, but transitioning from a stressed-timed to a syllable-timed language might represent a much greater challenge. For example, the transition between English and French can be seen as maximally difficult, as it involves reliance on different acoustic cues and a different strategy (locating word offsets instead of word onsets.) This difficulty has already been reported for English-French bilinguals who did not adopt and apply the language-specific strategy of their non-dominant language and were shown to rely instead on some other domain-general strategies [10, 11].

Yet we know that the bilingual experience varies a great deal, even within a bilingual community. So one might ask if such a pattern of relying on domain-general strategies applies across all bilinguals, regardless of language proficiency, dominance, or everyday use. To investigate this question, we designed a cross-modal priming experiment in which English-French bilinguals (English L1) listened to English and French sentences containing ambiguous syllable strings, and had to determine if a picture prompt appearing at the ambiguous point was related to the heard sentence or not. If individual differences in the bilingual experience do have an impact on L2 word segmentation, we should not only observe differences in the segmentation behavior across languages, but also within the L2, as a function of these individual differences.

## 2. Method

### 2.1. Participants

Ten English-dominant, English-French bilingual participants were recruited from the Montréal area (aged 22 to 31 years; average 26.7 years; 2 men). All were right handed and reported normal hearing, normal or corrected to normal vision and no history of neurological or psychiatric disorders. Prior to the experiment, participants filled in a health and language history questionnaire covering extensive language proficiency, background and use information (modified from [12]). Additionally, an objective L2 proficiency estimate was acquired, based on the sentence repetition sub-task of the Clinical Evaluation of Language Fundamentals 3 (CELF-3, [13]). In this task, participants simply have to repeat heard sentences of various lengths, structures and complexity and a score is assigned to each trial with regard to the number of errors made by the speaker (from 3 points for correct repetition to 0 points for more than 3 mistakes). Each participant performed the task in English (L1) and French (L2) and an index was computed by comparing the results in the two languages (French score subtracted from English score. A score of 0 would indicate that the speaker is as proficient in both languages; the

 doi: 10.21437/SpeechProsody.2016-8

larger the difference, the greater the proficiency difference between L1 and L2). Although this task is designed to evaluate the language skills of children, we selected it to take into account the potentially low level of proficiency of our L2 speakers. All participants signed a written consent and were compensated for their time.

### 2.2. Stimuli

#### *2.2.1. Stimuli design*

For each language, a set of 40 sentence pairs and 80 picture prompts were designed. Each sentence pair was constructed around an ambiguous syllable string interpretable as either two monosyllabic words (e.g. Engl.: *key we*; Fre.: *but fait*) or one bisyllabic word (e.g. Engl.: *kiwi*; Fre.: *buffet*). The wording of the sentences preceding the ambiguous string was identical. The ending of one sentence was compatible with the monosyllabic interpretation and the ending of the other was compatible with the bisyllabic interpretation. For example:

(1) If you would like a *key we* can duplicate one.
(2) If you would like a *kiwi* I will buy one tomorrow.

Three picture prompts were then selected for each sentence pair. The pictures represented either the first monosyllabic word (e.g. Engl.: *a key*; Fre.: *a goal*), the bisyllabic word (e.g. Engl.: *a kiwi*; Fre.: *a buffet*), or an unrelated word (e.g. Engl.: *a basketball*; Fre.: *a car*).

#### *2.2.2. Visual stimuli*

To ensure the pictures represented the intended concepts, we asked native speakers to judge them on a 1-7 Likert scale (1=prototypical representation of the word, 7= nothing to do with the word). The pictures included in the experiment were judged by at least 13 raters and received an average score of 2.5 or lower.

#### *2.2.3. Auditory stimuli*

The sentences were recorded by a simultaneous French-English bilingual in a sound-attenuating booth. Signals were digitized using a digital recorder (Marantz, Model PMD670) with a sampling rate of 44.1 kHz, 16 bits. Auditory sentences were produced with context-specific natural prosody. Amplitude was normalized and the latency of the time point between ambiguous syllables was measured using GoldWave (version 6.15). Distinguishability between contexts was established by 5 native speakers of each language.

### 2.3. Procedure

Presentation of the audio-visual stimuli was blocked by language and counterbalanced across subjects such that half of the participants heard the English (L1) trials first. Audio sentences were presented using insert earphones (*Eartone* 3A, EAR Auditory Systems) and pictures were displayed on a flat screen monitor placed 50cm away from the listeners. Participants were informed that pictures would appear while they heard the sentences and were instructed to determine, via a key press, as quickly as possible, whether the picture was related to the portion of the sentence presented so far. Sound files and pictures were played back via Inquisit 4 (Milisecond Software). The sounds were delivered at a constant intensity (peak levels of 67 dBA at the ears). The picture prompt was presented at the offset of the first syllable of the ambiguous string. Participants were presented with each sentence paired with each picture, permitting us to compare responses to the same pictures presented with different utterances. See Figure 1 for a schematic representation of the stimuli.

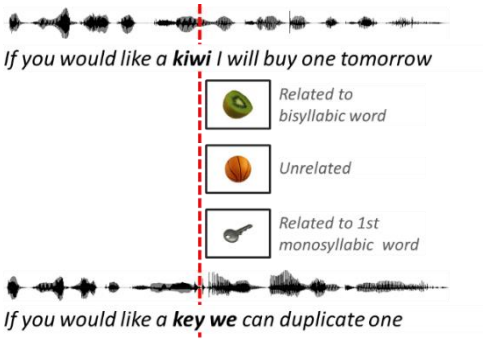


Figure 1: *Schematic representation of stimulus time-course of a pair of English sentences.*

## 3. Results / Discussion

### 3.1. Overall task effects

Figures 2 and 3 present the proportion of "related" responses given for each sentence condition (with two monosyllabic or one bisyllabic word in the ambiguous region) and each picture condition (illustrating the first monosyllabic word or the bisyllabic word) for English and French trials, respectively. Responses were considered regardless of response time. The "related" responses to the unrelated pictures were removed from the analyses since they were considered errors and represented less than 2% of trials. Overall, one can see that the pictures matching the sentence conditions yielded more "related" responses than pictures that mismatched the sentence version, although the pattern is more salient in English (L1) trials. That is, participants listening to their L2 (French) did not as consistently respond "related" when presented with sentences containing two monosyllabic words in the ambiguous region, along with a picture illustrating the first monosyllabic word; in fact, "related" responses did not differ from chance level in this condition.

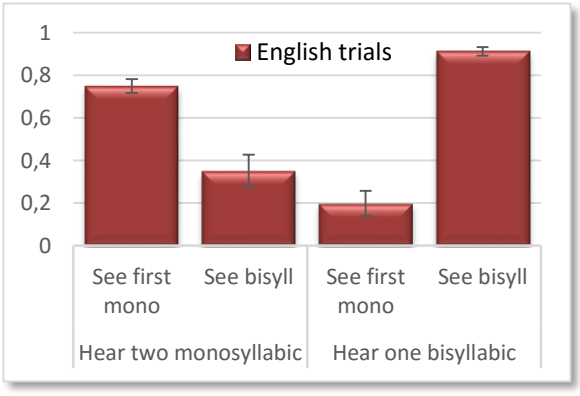


Figure 2: *Proportion of "related" response for English (L1) trials. Error bars indicate +/- 1 standard error of the mean.*

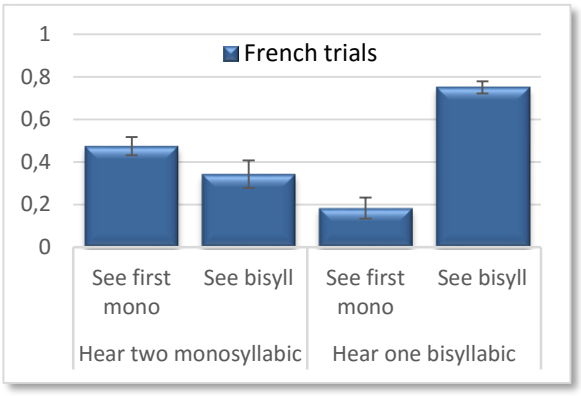


Figure 3: *Proportion of "related" response for French (L2) trials. Error bars indicate +/- 1 standard error of the mean.*

Statistical analyses were used to look for the effects of sentence and picture conditions in each language separately. Although we plan to use mixed-effects logistic regression to analyse the complete data set, preliminary results were analysed using 2 by 2 ANOVAs due to the limited number of participants. The analysis of English trials revealed a significant effect of picture condition [$F(1,10) = 27.033$, $p = .001$, $\eta^2 = .75$] and a significant interaction between sentence and picture conditions [$F(1,10) = 98.203$, $p < .001$, $\eta^2 = .916$]. Namely, the difference between picture conditions was greater when listening to sentences with a bisyllabic word in the ambiguous region compared to when the same pictures were presented during a sentence with two monosyllabic words. This difference may be due to the prosodic patterns available to listeners at the time of the decision. In both sentence conditions, the first syllable (whether a monosyllabic word or part of a bisyllabic word) is stressed and while listeners will interpret that syllable as a word onset, they may leave open the possibility that it will be followed by another syllable or represent a full monosyllabic word. In the case of bisyllabic words, upon presentation of the picture (at the offset of the initial stressed syllable), they will hear an unstressed syllable and reject the monosyllabic interpretation. In contrast, in the case of two monosyllabic words, upon presentation of the picture, they will hear a second stressed syllable, increasing the likelihood of interpretation of the initial syllable as a monosyllabic word, but not eliminating the multisyllabic interpretation, which would only be rejected fully upon presentation of the next syllable (which could still potentially comprise part of the initial word).

With regard to the French trials, the 2 by 2 ANOVA yielded significant effects for both sentence condition [$F(1,10) = 10.056$, $p = .011$, $\eta^2 = .528$] and picture condition [$F(1,10) = 78.311$, $p < .001$, $\eta^2 = .897$], as well as a significant interaction between the two factors [$F(1,10) = 35.855$, $p < .001$, $\eta^2 = .799$]. As for the English trials, we found a greater difference between picture conditions when presented during a sentence with one bisyllabic word in the ambiguous region. But unlike English trials, there was no clear difference between picture conditions when presented accompanying a sentence with two monosyllabic words in the ambiguous region. This finding was validated using a paired-sample t-test that yielded no significant difference between picture conditions within this (2-monosyllables) sentence condition [$t = 1.835$, $p = .1$, $\eta^2 = .272$]. So overall, our English L1 listeners were not able to distinguish between the two potential interpretations of the ambiguous region when it contained two monosyllabic words.

Taken together, these results indicate that sentences with two monosyllabic words in the ambiguous region are, not surprisingly, more prone to variation than sentences with one bisyllabic word in the same position, across languages. This may be due to the specific prosody of such sentences, as described above. This means that the decision process leading to the relatedness judgement is not as obvious for our English L1 listeners who might need to reanalyze the potential links between syllables before coming to a decision. Given the greater difficulty of this condition, we will focus our attention on this sub-group of trials to investigate the role of L2 proficiency on L2 word segmentation.

### 3.2. Effect of L2 proficiency on L2 word segmentation

To investigate the impact of L2 proficiency on L2 word segmentation, we divided our participants into two groups via a median split with regard to their L2 proficiency scores from the CELF-3 sentence repetition task. Figure 4 presents the proportion of "related" responses given when hearing French sentences with two monosyllabic words in the ambiguous region matched with both picture conditions (illustrating the first monosyllabic word or the bisyllabic word) with regard to relative L2 proficiency. Independent sample t-tests were used to observe the effect of L2 proficiency on picture relatedness judgments. The analyses revealed a significant difference across L2 proficiency groups only in the congruent condition, when listeners were presented with pictures illustrating the first monosyllabic word from the heard sentence [$t = -2,768$, $p < .05$, $\eta^2 = .657$], but not for the mismatching condition (when listeners were presented with pictures illustrating the bisyllabic word while listening to sentences with two monosyllabic words in the ambiguous region [$t = -.403$, $p > .1$, $\eta^2 = .039$]). Thus, listeners within the higher L2 proficiency group more often recognized the picture illustrating the first monosyllabic word of the ambiguous region as related to the heard sentence than did listeners from the lower L2 proficiency group. This indicates that listeners with higher L2 proficiency might have adapted their perception strategies to fit the prosody of French sentences, leading to more efficient segmentation processes even if French does not follow the same stressed-unstressed pattern found in their L1 (English).

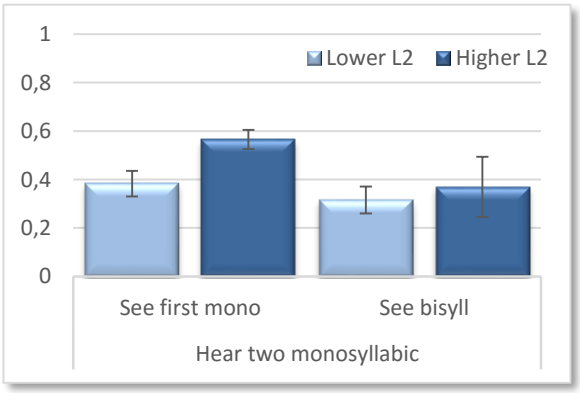


Figure 4: *Proportion of "related" response for French (L2) trials, with regard to estimated L2 proficiency. Error bars indicate +/- 1 standard error of the mean.*

## 4. Conclusions

Although these results represent preliminary data stemming from a limited pool of participants, we can already see that individual differences in bilingual experience seem to have an impact on L2 word segmentation. Here, we focused our attention on a single L2 proficiency index and found a significant difference across listeners with higher and lower L2 proficiency with regard to their processing of sentences containing segmentation ambiguities. These initial findings motivate further exploration of word segmentation in L2 listeners, focusing on individual differences across factors like age of acquisition, amount of L2 exposure and executive control capacities with a larger sample of participants.

## 5. Acknowledgements

We wish to thank Shanna Kousaie from the Montreal Neurological Institute for her help with participant recruitment and test design, and the team of testers of the Montreal Bilingual Brain Initiative: Don, Ally, Camille, Kristina, Yang, Nina and Emilia. This work was funded by an FRQSC team grant and the Blema and Arnold Steinberg Family Foundation.

## 6. References

1. A. Cutler, D. Dahan, and W. van Donselaar, "Prosody in the Comprehension of Spoken Language: A Literature Review". *Language and Speech*, vol. 40, no. 2, p. 141-201, 1997.
2. A. Cutler and T. Otake, "Rhythmic categories in spoken-word recognition". *Journal of Memory and Language*, vol. 46, no., p. 296-322, 2002.
3. P.W. Jusczyk, "How infants begin to extract words from speech". *Trends in Cognitive Sciences*, vol. 3, no. 9, p. 323-328, 1999.
4. P.W. Jusczyk, D.M. Houston, and M. Newsome, "The Beginnings of Word Segmentation in English-Learning Infants". *Cognitive Psychology*, vol. 39, no. 3–4, p. 159-207, 1999.
5. S.L. Mattys, et al., "Phonotactic and prosodic effects on word segmentation in infants". *Cognitive Psychology*, vol. 38, no., p. 465-494, 1999.
6. A. Christophe, et al., "Discovering words in the continuous speech stream: The role of prosody". *Journal of Phonetics*, vol. 31, no., p. 585-598, 2003.
7. A. Christophe, J. Mehler, and N. Sebastián-Gallés, "Perception of Prosodic Boundary Correlates by Newborn Infants". *Infancy*, vol. 2, no. 3, p. 385-394, 2001.
8. A. Christophe, et al., "Bootstrapping lexical and syntactic acquisition.". *Language and Speech*, vol. 51, no., p. 61-75, 2008.
9. A. Christophe, et al., "Phonological phrase boundaries constrain lexical access I. Adult data". *Journal of Memory and Language*, vol. 51, no., p. 523-547, 2004.
10. A. Cutler, et al., "The monolingual nature of speech segmentation by bilinguals". *Cognitive Psychology*, vol. 24, no., p. 381-410, 1992.
11. A.D. Endress and M.D. Hauser, "Word segmentation with universal prosodic cues". *Cognitive Psychology*, vol. 61, no. 2, p. 177-199, 2010.
12. P. Li, et al., "Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool.". *Bilingualism: Language and Cognition*, vol. 17, no. 3, p. 673-680, 2013.
13. E. Semel, E.H. Wiig, and W.A. Secord, *Clinical Evaluation of Language Fundamentals 3 (CELF-3).* . San Antonio, TX.: The Psychological Corporation, 1995.