Word segmentation in Czech

Václav Jonáš Podlipský[1], Nikola Paillereau[2,3], Kateřina Chládková[2,3] , Šárka Šimáčková[1]
[1] Palacký University Olomouc, [2] Czech Academy of Sciences, [3] Charles University
vaclav.j.podlipsky@upol.cz

How do we learn to parse uninterrupted streams of speech into words? Some have proposed that word segmentation is initially driven by speech prosody, predicting infants learning stress-timed languages to segment speech into stress units and infants learning syllable-timed languages to segment individual syllables [1]. However, others have shown that when segmenting speech, infants first rely on statistical cues, such as the syllable-to-syllable transition probabilities (TPs), higher within and lower between words [2–6]. The question of how word segmentation develops is not trivial, partly because to this date most research has focused on learners acquiring English, German or Dutch, and French or Spanish, which are exemplars of the so-called stress- and syllable-timed rhythm classes, respectively.

To understand better the contribution of prosodic and statistical cues to word segmentation, it is also needed to study learners from languages not falling neatly into one of the putative rhythm classes. Czech is one such language [7]. What is more, the Bohemian Czech (BCz) variety has stress fixed on the word-initial syllable whereas Silesian Czech (SCz) on the penultimate syllable, a difference the present study takes advantage of (fixed-stress languages are also understudied). We prepared artificial languages (i) with statistical cues only, (ii) with statistical and congruent prosodic cues (raised pitch and longer vowels in either initial or penultimate syllables), and (iii) with prosodic cues only. Thus we will be able to test if infants acquiring BCz and SCz segment words better with both statistical and prosodic stress cues and if segmentation is influenced by the native variety.

Before running the study with infants, it was necessary to assess how Czech adults segment words and if they display any variety-specific reliance on prosody. Using an online experiment, data were collected from 333 BCz and 158 SCz Czech adults. (Additional 109 Moravian Czechs took part but their results are not discussed here.) The volunteers heard 3.3-minute streams of speech (naturally-produced, edited) in the artificial languages (A and B, with swapped identities of statistical words and part-words, near-replicating the languages used by [4]) and then completed a forced-choice post-test in which they heard pairs of syllable triplets and decided which one was more likely a word in the language they had heard.

A generalized linear mixed model for exposure languages containing statistical cues without and with initial or penultimate prosodic cues detected an interaction of Prosody, listeners' native Variety, and exposed Language version (est = -0.271, se = 0.134, t = -2.021, p = .043), see Fig. 1. An analysis of exposure languages with prosodic cues only suggests a trend in native Variety affecting listeners' reliance on Initial versus Final stress cues (est = -0.623, se = 0.373, t = -1.670, p = .095), see Fig. 2. The results show that for TP-only exposure, BCz adults' word identification was at chance in both language versions (Fig. 1); as it was e.g. in [8]. With word-initial prosodic cues added, BCz adults identified words better, as expected, but only in version A. For prosody-only exposure, BCz adults preferred initial over final prominence parsing (Fig. 2), again as expectable. Next, with TPs only, SCz adults identified words above chance, and, like BCz listeners, they also identified words better with added word-initial than penultimate stress (Fig. 1). Thus, for the SCz adults penultimate prominence never enhanced word identification, which was unexpected. It seems the SCz adults exposed to a language with TPs and penultimate prominence (unlike SCz adults exposed to languages with prosodic cues only, cf. Fig. 2) tended to interpret the longer vowel as final lengthening.

The present results for Czech adults show that for a fixed-stress language falling outside the putative rhythm classes the role of prosody depends clearly on the stress placement and cue use in the native variety. Also, testing a so-far uninvestigated population, it turns out that adults do not necessarily demonstrate an ability to segment words based on syllable-to-syllable transition probabilities alone, at least with the methodology based on [4].

Infants from each native variety are now being tested. We will compare how they segment words after familiarization with TPs, prosodic cues, or both (aiming at n = 32 per variety and condition). Preliminary results on infants will also be presented at the conference.
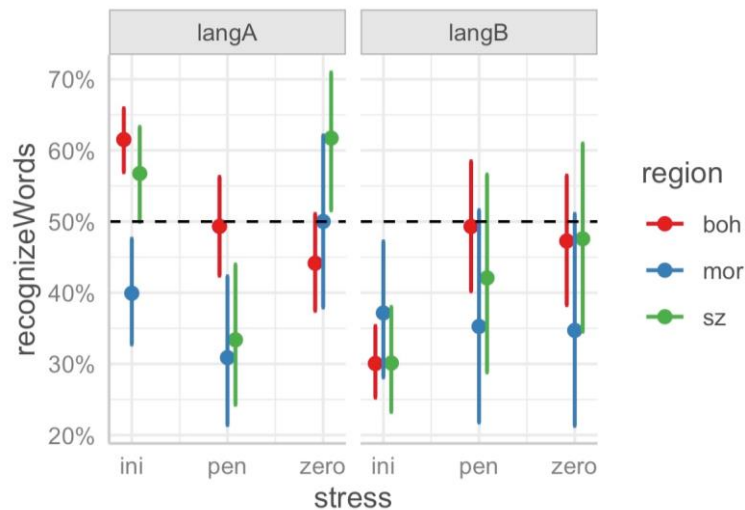
Figure 1. Model-predicted probabilities of choosing statistical words over part-words for languages with TPs: TPs as the only cue (zero), and with added initial (ini) or penultimate (pen) prosodic cues. Split by participants' native variety of Czech: Bohemian (boh), Silesian (sz), and Moravian (mor).
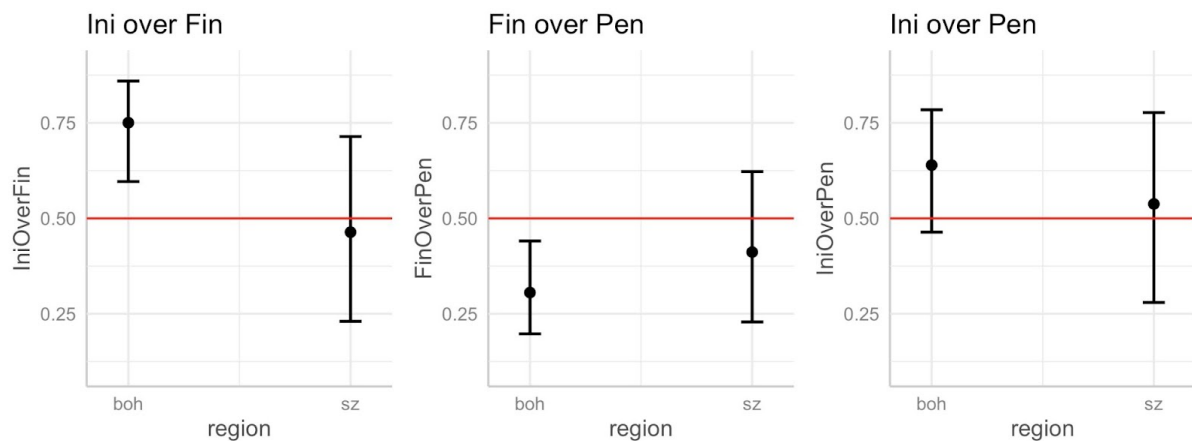

Figure 2. Modeled-predicted word-parsing preference split by participants' native variety (Bohemian [boh], Silesian [sz], and Moravian [mor] Czech) for participants exposed to languages with prosodic cues only.

References
[1] Nazzi et al., *J. Mem. Lang.*, 2006
[2] Jusczyk & Aslin, *Cogn. Psychol.*, 1995
[3] Saffran et al., *Science*, 1996
[4] Aslin et al., *Psychol. Sci.*, 1998
[5] Thiessen & Saffran, *Dev. Psychol.*, 2003
[6] Thiessen & Saffran, *Percept. Psychophys.*, 2004
[7] Dankovičová, J., & Dellwo, V., *Proc. of 16th ICPhS*, 2007
[8] Dal Ben et al., *Mem. Cogn.*, 2021