Rethinking Prosodic Phrasing

Fernanda Ferreira, Julie Bannon, Madison Barker, Beverly T. Cotter, Casey D. Felton, Barbora Hlachova, and Adrian Zhou (University of California, Davis)

Theories of processing, syntax, and even prosody have been heavily influenced by written language and its conventions. This is problematic because spoken and written language differ in major ways. For example, speakers make errors and correct themselves, and people talking spontaneously tend to produce shorter utterances with simpler syntactic structures [1]. Much of what we know about prosody, whether it be the features speakers are thought to generate or the ones listeners are believed to process, are based on forms that are atypical in spoken language. In addition, competence-based approaches that posit a syntactic or prosodic constituency from which prosodic features such as duration, intensity, and pitch are assumed to derive [2, 3] fail to capture performance-based effects [4] related to working memory and breathing constraints, as well as factors associated with planning and recovery.

However, recent developments in data analysis methods allow for the study of unrehearsed and/or spontaneous spoken language [5]. Corpora of naturally produced speech can be syntactically and prosodically analyzed, dramatically expanding the amount of data available for developing theories of prosody and freeing researchers from the need to construct bespoke items for theory testing. Our lab has recently embarked on a research program to characterize naturalistic prosody using these tools. The goal is to develop more ecologically valid theories of prosody in language production and thereby better capture the prosodic features that speakers produce and that listeners would typically have access to.

The utterances we have analyzed are unrehearsed descriptions of real-world scenes. Subjects (n=30 per study) freely describe 30 scenes for 30 seconds. The speech is then transcribed using Whisper, parsed using an automatic syntax parser (spaCy), and prosodically analyzed using the Wavelet Prosody Toolkit [5] for the presence of acoustic features associated with prosodic boundaries: word lengthening, pausing, and intensity and pitch variations. Each word is assigned a numerical value corresponding to the number of syntactic brackets around it and another number quantifying the strength of the prosodic boundary on/after the word, ranging from 0 (no boundary) to ~3 (strong boundary) [6]. In future studies we will apply the same analyses to other types of one-way speech as well as interactive language.

Figure 1 shows one set of surprising findings. Across four different studies, averaging over subjects and scenes, we display the correspondence between syntactic and prosodic boundaries, with both boundary types dichotomized for data presentation purposes. In each case, almost half the words show a mismatch between prosody and syntax: a strong syntactic boundary is associated with a weak prosodic boundary, and vice versa. In other words, contrary to the assumptions of most competence-based theories of prosody as well as most studies of prosody that require speakers to read prepared stimuli, there is only a moderate alignment between prosodic and syntactic structure. Figure 2 shows the distribution of prosodic boundaries of various strengths, for the same four studies. The findings show a continuous range of prosodic boundary strengths, peaking at two values: either no boundary or a relatively mild boundary (around 1 on a scale of 0-3), with strong prosodic boundaries being quite rare. Different speech styles (e.g., descriptions of events vs. states, speaker age) also influence the distribution of prosodic boundary strengths.

We suggest these results challenge some aspects of categorical approaches to prosody. We propose an alternative, continuous approach according to which, after each word, some number of prosodic features may be generated, ranging from none to a large enough set to constitute a strong prosodic boundary. The factors that determine the number of prosodic features generated reflect the production system’s dynamic need to plan and recover as an utterance unfolds. In addition, as recent work shows that the most diagnostic cue to a prosodic boundary is speech rate (pre-boundary lengthening and phrase-initial acceleration;[7]), we hypothesize that many properties of prosodic phrasing are attributable to differences in resource availability at the beginnings and ends of production units.

Figure 1: Percentage of words in a 30-second description with either a closing syntactic phrase boundary (light blue bars) or no phrase boundary (dark blue bars), plotted against presence of a weak or strong prosodic boundary (defined as in Figure 2), across four studies. The inner two bars in each figure are the cases in which the strength of prosodic and syntactic boundaries do not align. Panel A shows data for a study in which young adult speakers (18+ yrs) described actions that could be undertaken in the scene; B shows simple scene descriptions produced by older adults (65+ yrs); C shows simple scene descriptions produced by young adults; and D is the same as C except that a deadline to initiate the scene description was imposed. Data are averaged over the 30 subjects and scenes in each study.

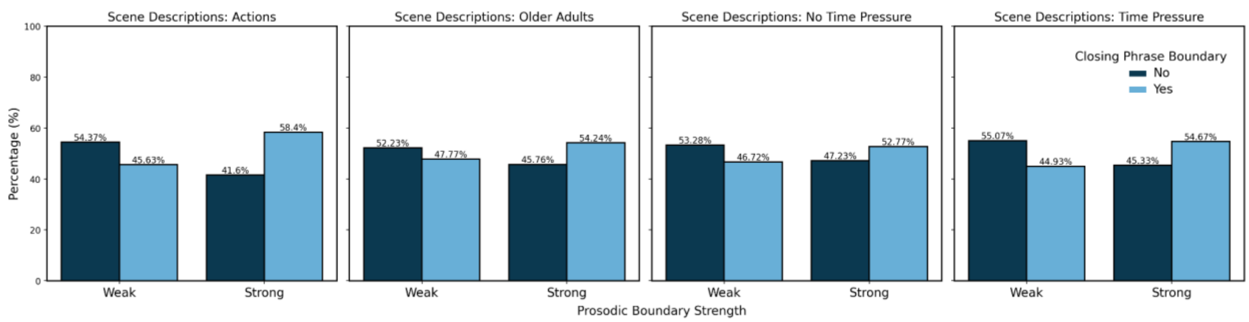
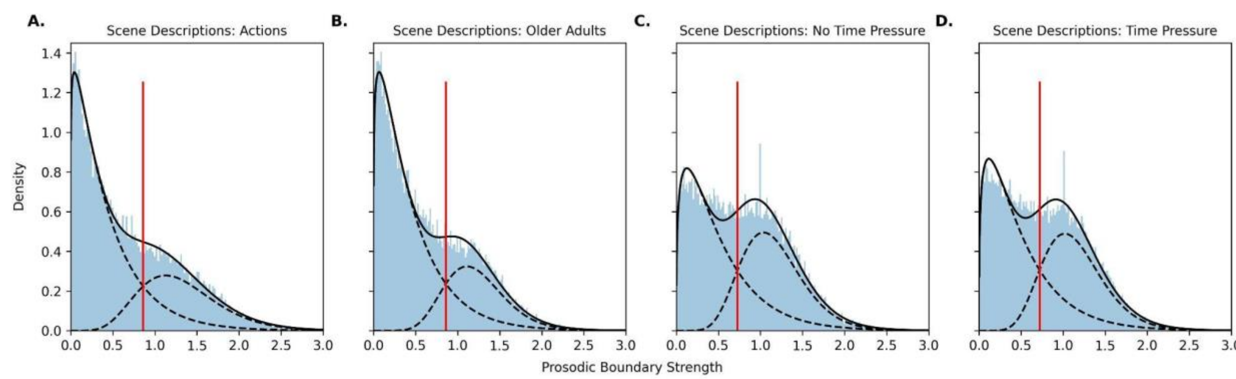


Figure 2: Distribution of prosodic boundary strengths of individual words as calculated by the Prosody Wavelet Toolkit using methods described in [5]. As in [5], we modeled our data using gamma mixture models to identify two distributions in the data set. The red line is the crossover point of the two distributions and represents the cutoff for weak vs. strong prosodic boundaries.



References

[1] Corps, R. E., Knudsen, B. & Meyer, A. S. *Cognition* **223**, 105037 (2022).
[2] Selkirk, E. *Phonol. yearb.* **3**, 371–405 (1986).
[3] Ferreira, F. *Language and Cognitive Processes* **22**, 1151–1177 (2007).
[4] Wagner, M. & Watson, D. G. *Language and Cognitive Processes* ***25***, 905–945 (2010).
[5] Suni, A., Šimko, J., Aalto, D. & Vainio, M. *Computer Speech & Language* ***45***, 123–136 (2017). Wavelet Prosody Toolkit. https://github.com/asuni/wavelet_prosody_toolkit
[6] Degano, G. *et al*. *Commun Biol* **7**, 1–10 (2024).
[7] Biron, T. *et al. PLOS ONE* **16**, e0250969 (2021).