Syllable prominence and prosodic phrasing in spoken prose

Isabelle Franz[1,2], Christine Knoop[1], Gerrit Kentner[1,2], Sascha Rothbart[1], Vanessa Kegel[1], Julia Vasilieva[1], Sanja Methner[1] & Winfried Menninghaus[1]

[1]Max-Planck-Institute for Empirical Aesthetics, [2]Goethe University Frankfurt

Metrical grids are supposed to reflect relative syllable prominence (Liberman & Prince, 1977), and partly account for the domains of the Prosodic Hierarchy (Halle & Vergnaud, 1987). However, their use for empirical studies is limited to highly controlled and short sentences. Also, current systems using metrical grids for syllable prominence prediction focus on decoding small verses (for poetry see Lerdahl, 2001), or on syntax/semantic-based automatic decoding of sentences that need to be annotated syntactically (Windmann et al., 2011). A replicable system for manually coding syllable prominence and prosodic boundaries in longer sentences or even texts is lacking so far, let alone its validation with the phonetic realization.

Based on work in the fields of metrical phonology (Kiparsky, 1966; Liberman & Prince, 1977) and existing prominence and pause coding systems (Gee and Grosjean 1983; Windmann et al., 2011), we developed a manual for coding syllable prominence (yielding up to 9 degrees of prominence) and prosodic boundaries (with 6 degrees of juncture). Figure (1) shows the basic workings of the system with prominence and boundary strength determining each other. The manual consists of a set of rules that are to be applied in a prescribed order; these rules mainly refer to simple cues in the text, like word/syllable count, part of speech, word position and punctuation.

Three independent annotators applied the coding system to the beginning pages of four different German novels (~90 000 syllables). With an inter-annotator agreement close to 1 (Cohen's κ .90 - .96), the conflicting cases were discussed and solved between the annotators resulting in a final consensus coding. We used the consensus coding to predict relative syllable prominence and prosodic boundary strength in the phonetic realization. As for syllable prominence, we hypothesized that a higher number of beats correlates with greater syllable duration and F0 range in the phonetic signal. As for prosodic boundaries, we assume a correlation of predicted prosodic boundaries strength with longer pauses in the phonetic realization. For the validation of the coding system eight professional speakers read the text described above aloud. We annotated the speech signal automatically, using MAUS (Schiel, 1999), matching the spoken syllables with citation form syllables. Using PRAAT (Boersma & Weenink, 2019), we extracted duration and F0 range for each syllable. These parameters were compared to predicted syllable prominence and prosodic boundary strength.

The results are shown in Figure (2-4). The validation with the speech signal and the high interrater agreement show that our annotation system predicts syllable prominence and prosodic boundaries to a highly reliable and replicable degree. Hence, we present the first validated annotation system for decoding the phonetic elements of prose rhythm. Our future work will focus on applying the system to a larger corpus of text, probably in a partly automatized process. There are numerous potential applications of the coding system, covering author profiling and style recognition, synthetic speech, and (psycho)linguistic research on prosody.

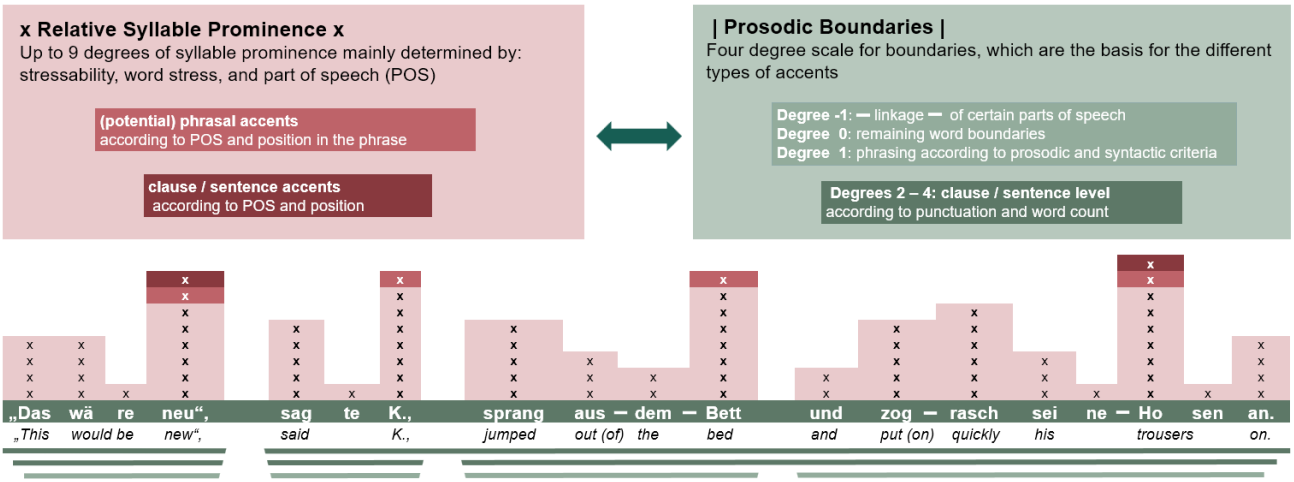

Figure 1: Basic workings of the coding system (upper panel) and an example sentence (lower panel). The metrical grid above the example sentence shows the prominence level of each syllable (in pink), the green lines below show the degree of the prosodic boundaries.

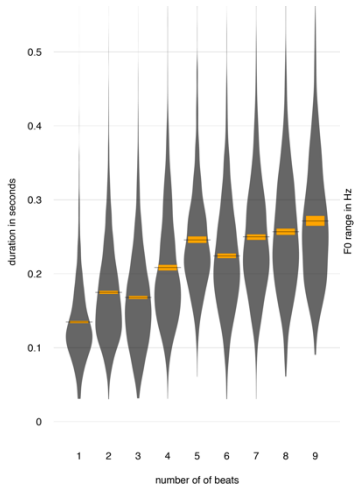

Figure 2: Syllable duration (in sec) by predicted syllable prominence (number of beats). The yellow bar shows the CI for the mean.

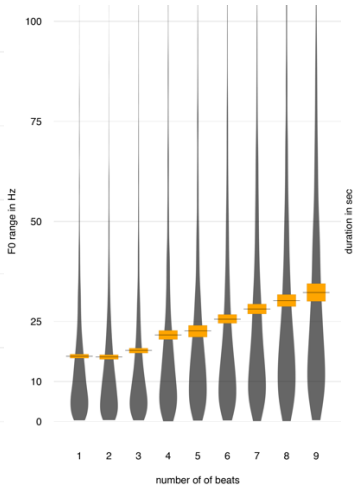

Figure 3: F0 range (in Hz) by predicted syllable prominence (number of beats). The yellow bar shows the CI for the mean.

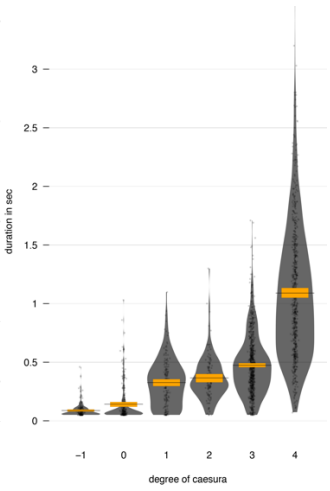

Figure 4: Pause duration (in sec) by predicted strength of prosodic boundaries (scale from -1 to 4). The yellow bar shows the CI for the mean.

References

Boersma, P., & D. Weenink (2019). *Praat: doing phonetics by computer* (Version 6.0. 52)[Windows].
Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. Cognitive psychology, 15(4), 411-458.
Halle, M., & Vergnaud, J. R. (1987). Stress and the cycle. *Linguistic Inquiry*, *18*(1), 45-84.
Kiparsky, P. (1966). Über den deutschen Akzent. Studia grammatica, 7, 69-98.
Lerdahl, F. (2001). The sounds of poetry viewed as music. *Annals of the New York Academy of Sciences, 930*(1),337–354.
Liberman, M., & A. Prince (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*(2): 249–336.
Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. *14th International Conference of Phonetic Science*, 1–7 August 1999, San Francisco.
Windmann, A., I. Jauk, F. Tamburini, & P. Wagner. (2011). Prominence- based prosody prediction for unit selection speech synthesis. *12th Annual Conference of the International Speech Communication Association*, 27–31 August 2011, Florence.