Bridging the Gap: Exploring a Middle-Way Approach for Prosodic Annotation

Aleksandra Ćwiek[1], Alina Gregori[2], Paula G. Sánchez-Ramón[2,3], Frank Kügler[2] & Pilar Prieto[4,3]

[1]Leibniz-Centre General Linguistics, [2]Goethe University Frankfurt, [3]Universitat Pompeu Fabra, [4]ICREA

cwiek@leibniz-zas.de

Prosodic annotation is crucial for linguistic research, allowing in-depth exploration of communication subtleties across languages and contexts. Manual methods of prosodic annotation are recognized, however, they pose challenges for prosody research. Multiple annotators' agreement varies significantly (even between 60% and 90%, Breen et al., 2012), leading to potential errors and manipulation. Time intensive manual annotation limits data availability, impeding reliable computational models for comprehensive and robust prosodic annotation across schemes and languages (e.g., Ananthakrishnan & Narayanan, 2008).

Our project adopts a middle-way approach, harnessing the expertise of skilled annotators and powerful computational tools while being cautious of challenges such as the complexity of prosody, interpretation of results, and ethical considerations regarding biases. We aim to establish connections between acoustic prominence cues and automatic predictions of perceived prominence. We analyze data from German and Catalan speakers, with perceived prominence annotated in target syllables on a scale from 0 to 3 (DIMA, Kügler et al., 2015). German and Catalan, with distinct prosodic properties (Krahmer & Swerts, 2007), facilitate cross-language computational exploration. We extract acoustic markers (like $f0$, intensity, duration) of prominence from target syllables, and use them as predictors in a Bayesian ordinal model. Our goal is to identify the most predictive features of prominence in prosody for German and Catalan, effectively bridging the gap between manual and computer-aided annotation. By establishing these links, we aim to address the challenges posed by tiresome manual annotation. Subsequently, we plan to employ our findings to build a classifier for automatic prominence assignment to utterances, which will undergo verification by human annotators.

References: Ananthakrishnan, S., & Narayanan, S. S. (2008). Automatic Prosodic Event Detection Using Acoustic, Lexical, and Syntactic Evidence. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(1), 216–228. • Breen, M., Dilley, L. C., Kraemer, J. & Gibson, E. (2012). Inter-transcriber reliability for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguistics and Linguistic Theory, 8(2)*, 277-312. • Krahmer, E. & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *J. Memory and Language 57*. 396-414. • Kügler, F., Smolibocki, B., Arnold, D., Baumann, S., Braun, B., Grice, M., Jannedy, S., Michalsky, J., Niebuhr, O., & Peters, J. (2015). DIMA: Annotation guidelines for German intonation. *Proc.18th ICPhS*, 317–320.