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

Author[1], Author[2]

[1] Affiliation [2] Affiliation

Prosodic annotation plays a critical role in linguistic research, enabling a detailed exploration of communication subtleties in diverse languages and contexts. Various methods for prosodic annotation exist, with some well-known and accepted, such as ToBI (Silverman et al., 1992). However, prosody research encounters challenges in the manual annotation of data, primarily regarding inter-rater reliability, where multiple annotators' agreement varies significantly (even between 60% and 90%, Breen et al., 2012), leading to potential errors and manipulation.

Manual annotation remains time-consuming and labor-intensive, resulting in limited data availability, hindering the development of reliable computational models for comprehensive and robust prosodic annotation across various schemes and languages (see Ananthakrishnan & Narayanan, 2008; Rosenberg, 2010 as examples for ToBI on Standard American English). To tackle this challenge, our project aims to adopt a middle-way approach, harnessing the expertise of skilled annotators alongside the power of computational tools while being cautious of potential challenges such as the complexity of prosody, interpretation of results, and ethical considerations regarding biases in training data and social impacts. By establishing robust computational models, we seek to significantly speed up the annotation process, generalize findings to new linguistic data, and pave the way for more extensive cross-linguistic studies in prosodic research.

In our project, we aim to establish connections between prosodic prominence marking and automatic predictions of focus conditions – focus representing a pragmatic domain of prominence (cf. e.g., Krifka, 2008). To achieve this, we will analyze data from German and Catalan speakers producing focus types as degrees of prominence in a semi-controlled environment. The perceived prominence of focus types in these data was annotated on a scale from 0 to 3 for prosodic prominence (DIMA, Kügler et al., 2015, 2019, 2022). German and Catalan are suitable languages for investigating the applicability of computational systems across different languages, given their distinct prosodic prominence marking in terms of rhythm class and accentuation patterns (cf. Krahmer & Swerts, 2007 for Germanic and Romance languages; Cole et al., 2019 for differences between Spanish, French, and English).

To capture acoustic markers of prominence, such as F0 (max peak, mean, range), intensity (max peak, mean, range), and duration, we will extract these measures from the accented syllables of focused words. These acoustic markers will serve as predictors in a Bayesian ordinal model to rate prominence. Our models will account for language-specific variations in acoustic features of prominence between Catalan and German.

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 focus-type 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. https://doi.org/10.1109/TASL.2007.907570

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. https://doi.org/10.1515/cllt-2012-0011

Cole, J., Hualde, J. I., Smith, C. L., Eager, C., Mahrt, T., & Napoleão de Souza, R. (2019). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics*, *75*, 113–147. https://doi.org/10.1016/j.wocn.2019.05.002

Krahmer, E. & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. Journal of Memory and Language 57. 396-414. Elsevier.

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica, 55*(3). 243–276.

Kügler, F., Baumann, S., Andreeva, B., Braun, B., Grice, M., Neitsch, J., Niebuhr, O., Peters, J., Röhr, C. T., Schweitzer, A., & Wagner, P. (2019). Annotation of German Intonation: Dima Compared with Other Annotation Systems. *Proceedings of the 19th International Congress of Phonetic Sciences*, 1297–1301.

Kügler, F., Baumann, S., & Röhr, C. T. (2022). Deutsche Intonation, Modellierung und Annotation. In C. Schwarze & S. Grawunder (Eds.), *Transkription und Annotation gesprochener Sprache und multimodaler Interaktion: Konzepte, Probleme, Lösungen* (pp. 23–54). Narr Francke Attempto Verlag.

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. *Proceedings of the 18th International Congress of Phonetic Sciences*, 317–320. https://kops.uni-konstanz.de/handle/123456789/32841.

Rosenberg, A. (2010). AuToBI --- A Tool for Automatic ToBI Annotation. *INTERSPEECH 2010*. 10th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. *ISCA*. 867–870.