

***Axio* Speech Corpus: design, structure, and preliminary analyses**

Katarzyna Klessa & Bożena Niećko-Bukowska

Institute of Linguistics, Faculty of Modern Languages and Literatures, Adam Mickiewicz University in Poznań
Al. Niepodległości 4, Poznań, Poland
{klessa, bniecko}@amu.edu.pl

Abstract

This contribution provides a description of a newly created speech corpus in Polish (*Axio* Corpus) along with the results of preliminary evaluation and analyses of the corpus contents. According to the assumptions underlying the corpus design, its first applications should enable two kinds of content analysis: (1) investigation of the properties of semi-spontaneous evaluative utterances, especially from the perspective of the notion of responsibility (cf. Jonas, 1979/1996); (2) inspection of phonetic-acoustic correlates of word stress in Polish semi-spontaneous and read utterances. By aiming at these two different research goals within the design of one common corpus we wished to extend the potential of the resource usability in line with the contemporary guidelines for language resource design and development (e.g., Váradi et al., 2008). The Polish *Axio* Corpus includes recordings of 32 Polish native speakers collected through a two-stage procedure. The first stage comprised of three task-oriented dialogues aiming at eliciting semi-spontaneous conversations. The dialogue tasks involved discussing photographs of everyday life situations potentially encouraging reflections concerning responsibility in its various meanings. Additionally, as part of the final task, the speakers cooperatively designed a news-style summary of their discussion. The second stage of the procedure was based on sessions recorded approximately four weeks after the first stage. Here, the speakers were requested to read a phrase list designed based on the materials produced in the first stage. This way, the same lexical content was obtained in both read and semi-spontaneous speaking style (cf. the approach by e.g., Eriksson & Heldner, 2015). We will report on preliminary observations regarding selected acoustic-phonetic features of the collected data and characterize the news-style reports in the social context of responsibility.

Keywords: speech resources, reusability of tools and resources, interdisciplinary research

References

- Eriksson, A., Heldner, M. (2015). The acoustics of word stress in English as a function of stress level and speaking style. *Proceedings of INTERSPEECH 2015*, Dresden, 41-45.
- Jonas, H. (1979) *Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation*, Frankfurt am Main; (1996) *Zasada odpowiedzialności. Etyka dla cywilizacji technologicznej*, (transl. M. Klimowicz), Wyd. Palatyn, Kraków.
- Váradi, T., Krauwer, S., Wittenburg, P., Wynne, M., & Koskenniemi, K. (2008). CLARIN: Common Language Resources and Technology Infrastructure. *Proceedings of the 6th International Conference on Language Resources and Evaluation*, Marrakech.