Laura Giacomini (University of Innsbruck) & Fritz Kliche (University of Hildesheim)

Title

Data acquisition in PhraseBase: core method and exploratory case studies

Abstract

Our presentation focuses primarily on the current data acquisition method for the lexicographic and ontological components of the PhraseBase project (DiMuccio-Failla & Giacomini, 2022). The evolution of the method, which focuses on the collection of phraseological data from corpora and other selected sources (Giacomini & DiMuccio-Failla, 2019), will be illustrated.

Parallel to the application of this core method, a number of studies have been carried out to explore opportunities for automating some steps of the lexicographic process. In the talk we will mention Kliche and Giacomini's study (2024, forthcoming) on the verb *follow* using contextualized word embeddings for the disambiguation of word senses. Fritz Kliche and Jannis Nolte finally report on first experiments using zero-shot text classification (Yin et al., 2019) for the disambiguation of reading variants in a case study on *paint* and *draw*.

References:

DiMuccio-Failla, P. & Giacomini, L. (2022): A proposed microstructure for a new kind of active learner's dictionary. In: *Lexicographica 38 (2022)*, de Gruyter, 475-499

Giacomini, L. & DiMuccio-Failla, P. (2019): Investigating Semi-Automatic Procedures In Pattern-Based Lexicography. In: *Proceedings of eLex 2019 Conference. Electronic lexicography in the 21st century*, October 2019 - Sintra (PT)

Kliche, F. & Giacomini, L. (2024, forthcoming): Exploring BERT's contextualized word embeddings: a suitable method for a lexicography-oriented analysis of argument structures? In: Giacomini, L. & Piunno, V. (eds.): *Patterns of Meaning in Lexicography and Lexicology*, Lexicographica Series Maior, De Gruyter.

Yin, W, Hay, J., & Roth, D. (2019): Benchmarking zero-shot text classification: datasets, evaluation and entailment approach. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China. Association for Computational Linguistics.