Using Contextual Information for Deep-Level Morphological Analysis

Donnerstag
09.03.2017
10:30 – 11:15
B4 1, Foyer

Petra Steiner
Institut für Deutsche Sprache
steiner@ids-mannheim.de

For detailed semantic processing, e.g. in information retrieval or the building of terminology databases, the recognition of hierarchical structures is a prerequisite. But concerning these structures, many word forms are leading to ambiguous structure interpretations. Using ontologies of specific domains can be helpful, if available (see Bretschneider & Zillner 2015 for compound splitting). Information about general semantic similarities (e.g. Ziering et al. 2016) does not take into account the ambiguities of linguistic forms if drawn from large corpora. By contrast, the methodological framework of this investigation builds on the hypothesis that morphological analysis can be improved by the specific contextual information of the lexical items without necessitating an ontology or other semantic networks.

Steiner & Ruppenhofer (2015) and Steiner (2016) developed a method for building parts of morphological structures by using counts from a morphological database and a corpus for computing the weighting measures, thereby using a wide notion of context. The current approach uses more restricted context definitions, as it works with frequencies of smaller and more homogenous texts and corpora. We use a corpus of a small domain, gather information from wider and narrower contexts and show to what extent these can improve morphological analyses.

References: • Bretschneider, C. & S. Zillner (2015): Semantic Splitting of German Medical Compounds. In: Král, P. & Matoušek, V., eds. *Text, Speech, and Dialogue: Proceedings of the 18th International Conference, TSD 2015, Pilsen, Czech Republic, September 14–17, 2015*. Cham: Springer International Publishing. 207–215. • Steiner, P. (2016): *Kontextbasiertes morphologisches Parsing. Poster presentation at the 38rd Annual Conference of the German Linguistic Society (DGfS 2016), University of Konstanz, February 24–26, 2016*. • Steiner, P. & J. Ruppenhofer (2015): Growing Trees from Morphs: Towards Data-Driven Morphological Parsing. In: Fisseni, B., B. Schröder & T. Zesch, eds. *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology (GSCL 2015), University of Duisburg-Essen, Germany, September 30 – October 2, 2015*. 49–57. • Ziering, P., S. Müller & L. van der Plas. (2016): Top a Splitter: Using Distributional Semantics for Improving Compound Splitting. In: *Proceedings of the 12th Workshop on Multiword Expressions, Berlin, Germany, August 7–12, 2016*. 50–55.