Using Contextual Information for Deep-Level Morphological Analysis

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:**  

333