Parsing German with Verb Subcat Frames

Bich-Ngoc Do
Leibniz ScienceCampus
do@ cl.uni-heidelberg.de

Ines Rehbein
Leibniz ScienceCampus
rehbein@
cl.uni-heidelberg.de

Anette Frank
Leibniz ScienceCampus
frank@ cl.uni-heidelberg.de

Tag
Datum
Zeit
Raum

We present here our ongoing work on using verb subcategorisation frames to improve neural dependency parsing for German. Recent years have seen great advances in syntactic parsing, employing neural network techniques and automatically learned features via embeddings. Those systems can be trained to parse any languague, and parsing accuracy for English is high (Dozat & Manning, 2017). Results for parsing languages other than English, however, are often substantially lower since those universal parsing systems do not pay attention to the characteristics of each specific language.

German, for instance, has a much richer morphology than English, combined with a semi-free word order and case syncretism. The combination of those characteristics makes parsing a challenging task and results in higher errors rates especially for core verbal arguments. In this work, we try to improve a neural parser by pruning possible analyses based on subcat frame information.

We use the Subcat Frame Database (Scheible et al., 2013) and subcat frames extracted from the Tüba-D/Z corpus (Hinrichs et al., 2004) to build a neural classifier that predicts the frame for each verb in the sentence. Our baseline parser is a reimplementation of the *parsing as head selection* system (Zhang et al., 2017), a two-stage parser that builds an unlabelled tree for each sentence by selecting the most probable head for each word, and then assigns the dependency labels. We will integrate the frame classifier into the parsing system via *joint learning* or *multi-task learning* and report on the success of this approach.

References: ● T. Dozat & C. Manning (2017): Deep biaffine attention for neural dependency parsing. In *Proceedings of ICLR 2017.* ● E. Hinrichs, S. Kübler, K. Naumann, H. Telljohann & J. Trushkina (2004): Recent developments in linguistic annotations of the TüBa-D/Z treebank. In *Proceedings of TLT 2004*, 51–62. ● S. Scheible, S. Schulte im Walde, M. Weller & M. Kisselew (2013): A compact but linguistically detailed database for German verb subcategorisation relying on dependency parses from a web corpus: Tool, guidelines and resource. In *Proceedings of the 8th Web as Corpus Workshop*, 63–72. ● X. Zhang, J. Cheng & M. Lapata (2017): Dependency parsing as head selection. In *Proceedings of EACL 2017*, 665-676.