

Mittwoch  
08.03.2017  
15:45 – 16:30  
B4 1, Foyer

---

## How to make manual annotation more efficient - Ensemble Dependency Parsing as a preprocessing step to obtain high-quality annotation in an efficient way

---

Heike Zinsmeister  
Universität Hamburg

[Heike.Zinsmeister@uni-hamburg.de](mailto:Heike.Zinsmeister@uni-hamburg.de)

Jessica Sohl  
Universität Hamburg

[Jessica.Katharina.Sohl@studium.uni-hamburg.de](mailto:Jessica.Katharina.Sohl@studium.uni-hamburg.de)

The “multistrategy approach” of ensemble tagging is a method of tagging texts automatically by a cluster of taggers, resulting from training different learning algorithms on the same data (von Halteren et al. 2001, p. 201, e.g., Rehbein et al. 2014).

In this poster, we will present ongoing work of using this method for dependency parsing as a pre-processing step for manual annotation of German textbook texts. An ensemble of three parsers is used to annotate sentences automatically and to identify sentences that need to be checked manually. We deviate from the original ensemble idea by including a rule-based system in the cluster in addition to the machine learning approaches. We will present qualitative analyses of sentence types that the ensemble fails to parse correctly and provide cross-validation results for two training corpora (Hamburg Dependency Treebank (Menzel et al. 2014), TIGER Corpus v2.2 dependency conversion (Seeker & Kuhn 2012)) and the accuracy of adapting the trained ensemble to our target domain of textbook texts.

**References:** • Foth, Kilian, Köhn, Arne, Beuck, Niels, Menzel, Wolfgang (2014): Because Size Does Matter: The Hamburg Dependency Treebank. In: *Proceedings of the Language Resources and Evaluation Conference 2014 / European Language Resources Association*. • Rehbein, Ines, Schmalowski, Sören, Wiese, Heike (2014): The KiezDeutsch Korpus (KiDKo) Release 1.0. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation*, p. 3927-3934. Reykjavík, Iceland. • Seeker, Wolfgang, Kuhn, Jonas (2012): Making Ellipses Explicit in Dependency Conversion for a German Treebank. In: *Proceedings of the 8th International Conference on Language Resources and Evaluation*, p. 3132-3139, Istanbul, Turkey. • Van Halteren, Hans, Daelemans, Walter, Zavrel, Jakob (2001): Improving Accuracy in Word Class Tagging through the Combination of Machine Learning Systems. In: *Computational Linguistics* 27(2), p. 199-229.