

Universal Dependencies for Slavic Languages

Daniel Zeman

Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University
Malostranské náměstí 25, 118 00 Praha/Prague, Czech Republic
zeman@ufal.mff.cuni.cz

Universal Dependencies (UD) is an international project that seeks to define guidelines for annotation of morphology and syntax, applicable to all natural languages. Its motivation is to find a common lingua franca for people and tools that deal with annotated linguistic data. Besides guidelines, UD also releases dependency treebanks converted to the UD annotation style; this data is becoming an important resource for multilingual NLP research and applications. In my talk, I will give a brief introduction to UD in general, and then I will focus on phenomena specific to Slavic languages. The UD release 1.2 includes six Slavic languages and others are being worked on. This gives us plenty of material for comparative studies, that in turn can (and should) further contribute to improved cross-linguistic consistency and fine-grained UD guidelines for Slavic languages. I will present observations from the UD data, as well as some results of dependency parsers trained on the data.