

Harvest Time – Explorations of the Swedish Treebank

Joakim Nivre

Uppsala University

Work on building a large treebank for Swedish started at about the same time as the TLT workshop series and reached a significant milestone this year with the second release of the Swedish Treebank, a corpus developed by merging and harmonizing the existing corpora Talbanken and the Stockholm-Umeå Corpus. In this talk, I will first present the treebank itself, explaining how it was developed using cross-corpus harmonization and annotation projection and describing the final result, which is a multi-representation treebank including annotation of phrase structure, grammatical functions and dependency structure. I will go on to describe ongoing work at exploiting the treebank for parser development, using data-driven methods for dependency parsing, and I will end by discussing our plans to use the treebank for cross-framework parser evaluation, in particular for comparing constituency-based and dependency-based parsing methods.