

Title: Coreference resolution for Universal Dependencies

Author: Ján Faryad

Department: Institute of Formal and Applied Linguistics

Supervisor: Mgr. Michal Novák

Abstract:

Coreference is an important tool for maintaining of the text coherence. Up to now, there has been no possibility to mark it in Universal Dependencies (UD), which is a project for universal description of morphology and dependency syntax. This work presents a way how to mark coreference in the UD project. It also includes a conversion of data with coreference annotation from the corpora PDT 3.0 and OntoNotes 5.0 with using a tool UDPipe for an automatic analysis of text in the UD style.

This work is also aimed to implement a system for automatic resolution of pronoun coreference using machine learning. Finally, the quality of the system is evaluated by simple way. The design of the program emphasizes the language independence and compatibility with the Udapi interface, which is used for processing of the UD data.

Keywords: coreference resolution, coreference, anaphora, Universal Dependencies, UD