

Annotation of topic-focus articulation in a parallel English-Czech dependency treebank

- Aim: annotation scheme for Topic-Focus Articulation (TFA) within the parallel PCEDT and PEDT treebanks
 - English part – newly created annotation scheme
 - Czech part – revision of the recent annotation scheme for TFA
- Team
 - E. Hajičová: project guarantee
 - J. Straková: project leader, IT support
 - B. Štěpánková: research of conjunctions and particles
 - K. Veselá: Czech part of annotation manual
- Supported by: The Charles University Grant Agency (GAUK 158010)

Work on PDT, PEDT and PCEDT data

- browsing the trees and localization of potential problems of the recent annotation scheme
- analysis of query results and further research of the above-mentioned problems
- suggestions of their solution in a newly refined annotation scheme
- checking of relations of similar phenomena in Czech and English parallel trees
- comparison of characteristics of authentic Czech text (on PDT data) and translated Czech text (on PCEDT data)

Current progress on Czech part

- selected problems of annotation
 - position of focalizers in tectogrammatic structure
 - homonymy of particles
 - distribution of contrastive topics
 - definition of proxy focus and quasi focus
 - TFA in coordinated structures
- to-do
 - systemic ordering, deep word order in topic, non-projective constructions, idioms