Quality and Efficiency of Manual Annotation: Pre-annotation Bias
Marie Mikulová, Milan Straka, Jan Štěpánek, Barbora Štěpánková, Jan Hajíč
{mikulova,straka,stepanek,stepankova,hajic@ufal.mff.cuni.cz}
Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

Dependency Syntax Annotation
• within multi-layered treebank for Czech: Prague Dependency Treebank – Consolidated 2.0
• structure and syntactic function label (25 types): Predicate, Subject, Object, Attribute, Adverbial etc.
• plus additional features: ellipsis, parenthesis, member of coordination (25x23=200 types of labels)
• 2,000,000 tokens to be annotated

Annotation Support Tools

Experiment Design

Results

Conclusions
 Automatic pre-annotation
• efficient tool for manual syntactic annotation
• increase in speed and consistency of annotation
• no reduction in quality

Use of other support tools
• together they significantly increase the quality of annotation
• no increase in annotation time needed (much)

Data from experiment available: http://hdl.handle.net/11234/1-4647