K-Best Spanning Tree Dependency Parsing With Verb Valency Lexicon Reranking

Željko Agić
Department of Information and Communication Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
zagic@ffzg.hr

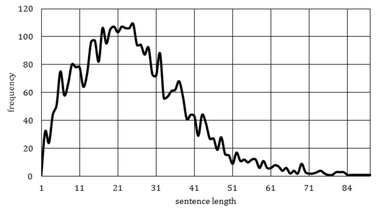
CESAR
CEntral and South-east europeAn Resources

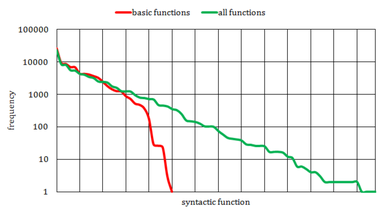
Introduction and setup

Depencency parsing of morphologically complex languages still poses a challenge, especially when paired with small dependency treebanks.

We aimed at improving dependency parsing accuracy for Croatian by using the Croatian Dependency Treebank (HOBS) and developed a novel generic dependency parsing approach.

The approach is tested on HOBS and compared to standard transition- and graph-based parsers.

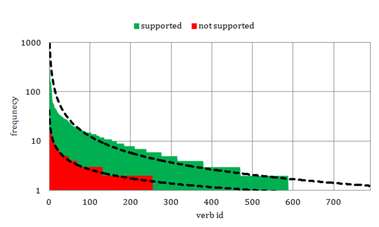


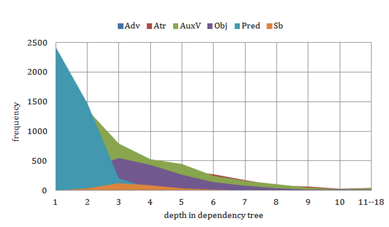


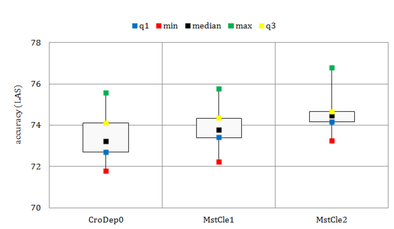
Suggested model

Notion of verb valency is inherent in the treebank, but verb valency lexicons such as CROVALLEX do exist and explicitly encode verb valency.

We have developed a parser that reranks dependency trees suggested by a MST-based parser. Reranking is done by using a verb valency lexicon.



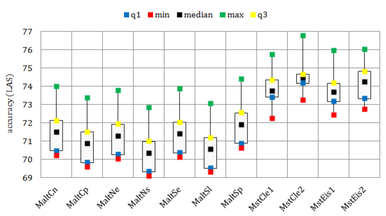


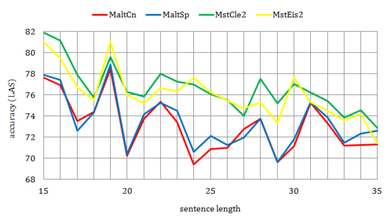


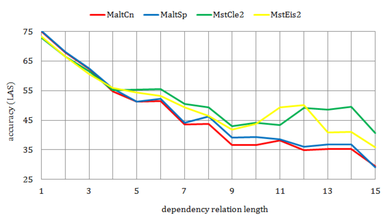
Existing parsers

MaltParser and MSTParser were used in eleven instances on HOBS with tenfold cross-validation. MSTParsers significantly outperformed MaltParsers.

The top-performing parser was a second-order arc-factored Chu-Liu/Edmonds parser from the MSTParser system (74.53% LAS, 81.69% UAS).







Parsing with CroDep

The model outperforms the existing parsers by at least 2.68% LAS in the same experiment setup on HOBS. Improvements are also observed on PDT.

The overall improvement is reached by a substantial increase in parsing accuracy for predicates, subjects and objects (more than 10.00% LAS each).