Syntactic Analysis of Natural Languages

Miloš Jakubíček

Natural Language Processing Centre
Faculty of Informatics, Masaryk University
Botanická 68a, 602 00 Brno, CZ

jak@fi.muni.cz

DTEDI, 7. 11. 2011

Contents

What Is Natural Language Processing (NLP)?

in terms of processing separate linguistic layers

- phonology/phonetics
- morphology
- syntax
- semantics
- pragmatics
- (logic)

in terms of NLP tasks

- information extraction/retrieval
- question answering
- summarization
- machine translation
- anaphora resolution
- named entity recognition
- speech synthesis/recognition
- computer lexicography
- ...

What Is Natural Language Processing (NLP)?

in terms of processing separate linguistic layers

- phonology/phonetics
- morphology
- syntax
- semantics
- pragmatics
- (logic)

in terms of NLP tasks

- information extraction/retrieval
- question answering
- summarization
- machine translation
- anaphora resolution
- named entity recognition
- speech synthesis/recognition
- computer lexicography
- . . .

Syntactic Analysis (Parsing)

- a well-known problem in Computer Science
- goal: to recover the structure of the input sentence
- result: usually some form of a parse tree
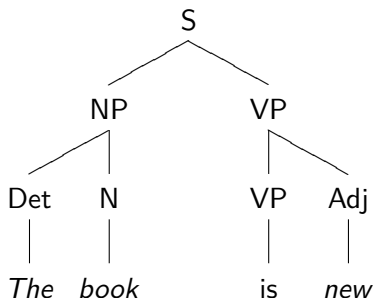
Parsing Methods

- rule-based: from a given grammar
- statistical: training on a syntactically annotated corpus (a treebank) using ML methods

Issues

- how to achieve high precision (language ambiguity)
- how to achieve wide coverage (language variety)
- how to measure parsing precision (correctness)
- how to achieve better applicability of results (interpretation)

What do papers about parsing say?
"*Parsing is a crucial step for many NLP applications.*"

What do people developing NLP applications say?
"*We tried to use a parser but it didn't improve the results of our application.*"

Aims of My Thesis

- to redefine parsing as a two-step problem:
    - what syntactic information do we need and in what format?
    - how to obtain it with high precision and wide coverage?
- elaborate on step 1 (theoretical part) with regard to:
    - practical applications of parsing
    - inter-annotator agreement on syntactic phenomena
    - descriptive adequacy of the format
    - inter-application usability of parsing
- develop a parser that will meet the requirements given in step 1 and step 2 (practical part) and evaluate it on particular applications

Achieved Results

on format of parsing results:

Mining Phrases from Syntactic Analysis (Jakubíček, Horák, Kovář, conference paper 2009)

Syntactic Analysis Using Finite Patterns: A New Parsing System for Czech (Kovář, Horák, Jakubíček, conference paper 2011)

on inter-annotator agreement in syntax:

Through Low-Cost Annotation to Reliable Parsing Evaluation (Grác, Jakubíček, Kovář, conference paper 2010)

on parsing precision:

Effective Parsing Using Competing CFG Rules (Jakubíček, conference paper 2011)

Full Morphosyntactic Analysis of Czech (Jakubíček, Horák, Šmerk, journal paper submitted 2011)

(publications indexed by Thomson Reuters listed only)

Bibliography

TBA :)