

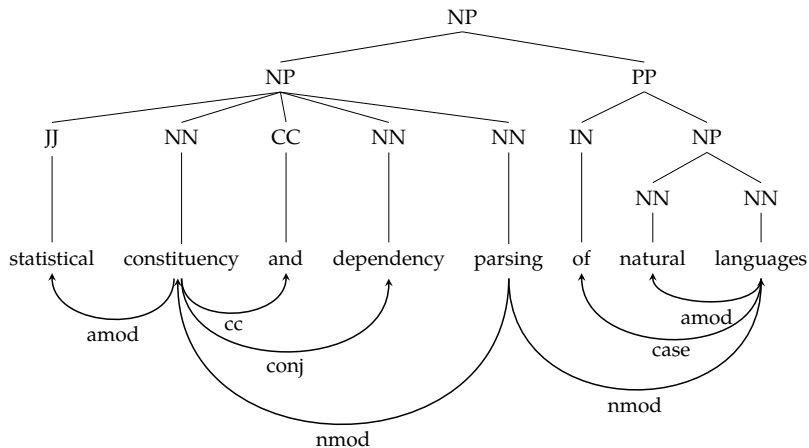
Statistical Parsing

Çağrı Çöltekin

University of Tübingen
Seminar für Sprachwissenschaft

October 25, 2016

This course is about ...



Ingredients of a parser

- ▶ A grammar

Ingredients of a parser

► A grammar

$S \rightarrow NP VP$	$N \rightarrow \textit{course}$
$NP \rightarrow NP PP$	$N \rightarrow \textit{parse}$
$NP \rightarrow A N$	$V \rightarrow \textit{parse}$
$VP \rightarrow V NP$	$A \rightarrow \textit{statistical}$
$PP \rightarrow P NP$	$P \rightarrow \textit{of}$

Ingredients of a parser

- ▶ A grammar

$$\begin{array}{ll}
 S \rightarrow NP VP & N \rightarrow \textit{course} \\
 NP \rightarrow NP PP & N \rightarrow \textit{parse} \\
 NP \rightarrow A N & V \rightarrow \textit{parse} \\
 VP \rightarrow V NP & A \rightarrow \textit{statistical} \\
 PP \rightarrow P NP & P \rightarrow \textit{of}
 \end{array}$$

- ▶ An algorithm for parsing

Ingredients of a parser

- ▶ A grammar

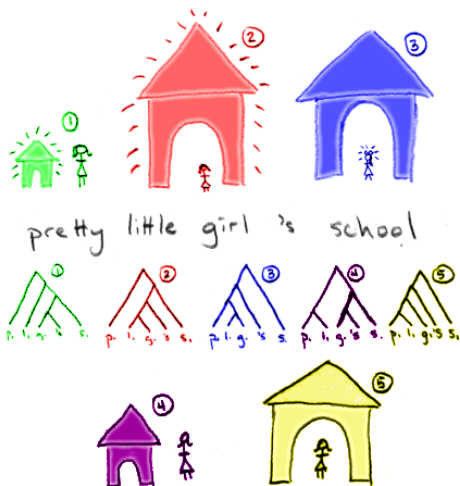
$$\begin{array}{ll}
 S \rightarrow NP VP & N \rightarrow \textit{course} \\
 NP \rightarrow NP PP & N \rightarrow \textit{parse} \\
 NP \rightarrow A N & V \rightarrow \textit{parse} \\
 VP \rightarrow V NP & A \rightarrow \textit{statistical} \\
 PP \rightarrow P NP & P \rightarrow \textit{of}
 \end{array}$$

- ▶ An algorithm for parsing
- ▶ A method for ambiguity resolution

We want our parsers to ...

- ▶ be efficient
Combinatorial nature of the parsing can easily result in intractability.
- ▶ be robust, but also not 'leaky'
 - ▶ We want the parsers not to fail easily, produce some useful output even if the sentence is not 'correct'
 - ▶ But we also do not want them to analyze incorrect sentences
- ▶ identify the most likely parse

Why do we need statistics/disambiguation?



Cartoon Theories of Linguistics, SpecGram Vol CLIII, No 4, 2008. <http://specgram.com/CLIII.4/school.gif>

Time/Place

- ▶ Time: Tue/Thu 12:00 – 14:00
- ▶ No textbook, but the following include useful basics
 - ▶ Daniel Jurafsky and James H. Martin (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. second. Pearson Prentice Hall. ISBN: 978-0-13-504196-3, esp. chapters 13–16
 - ▶ Sandra Kübler, Ryan McDonald, and Joakim Nivre (2009). *Dependency Parsing*. Synthesis lectures on human language technologies. Morgan & Claypool. ISBN: 9781598295962
- ▶ Office hours: Mondays 10:00–12:00, or email (ccoltekin@sfs.uni-tuebingen.de) for an appointment
- ▶ Course web page: <http://sfs.uni-tuebingen.de/~ccoltekin/courses/statistical-parsing/>

Course work

- ▶ Active participation is required
- ▶ Weekly/bi-weekly assignments, that will lead to implementation of a dependency parser at the end of the course
- ▶ Presenting/leading discussion on one parsing-related paper
- ▶ Reading the papers to be discussed in the class. You are also required to send two questions/discussion points about each paper via email
- ▶ A term paper describing the parser, experiments with multiple languages

The plan (tentative)

- ▶ Grammars, grammar formalisms
- ▶ Context-free parsing algorithms
- ▶ Statistical context-free parsing
- ▶ Dependency parsing
- ▶ Paper presentations

Your first assignment

Send 10 sentences in a language of your choice via email before the next lecture on Thursday.

- ▶ More 'unusual' the language is the better, it does not have to be your native language, a fair understanding is sufficient.
- ▶ Try to cover a range of 'interesting' syntactic phenomena, some ambiguity examples.
- ▶ Tip: grammar books often include such examples.
- ▶ Make sure they are 'real' sentences, avoid 'translationese'.
- ▶ Translations to English are welcome, but can also be later.

Next week

Grammars, grammar formalisms and treebanks.

- ▶ Phrase structure grammars, and their relation to automata
- ▶ Dependency grammars
- ▶ Brief notes on a few other grammar formalisms used in computational linguistics
- ▶ Treebanks

Bibliography



Jurafsky, Daniel and James H. Martin (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. second. Pearson Prentice Hall. ISBN: 978-0-13-504196-3.



Kübler, Sandra, Ryan McDonald, and Joakim Nivre (2009). *Dependency Parsing*. Synthesis lectures on human language technologies. Morgan & Claypool. ISBN: 9781598295962.