Formal Models in NLP: Parsing

Daniel Quernheim

April 17, 2012
What’s parsing?

- Syntactic analysis of a string
- CFG, CCG, Dependency Grammar

```
S
 /  \\  
NP VP
 /  \\
Kim V PP
 /  \\
lives P PN
 /  \\
in Australia
```
What’s parsing?

- Syntactic analysis of a string
- CFG, CCG, Dependency Grammar

```
\[
\begin{array}{cccc}
  \underline{Kim} & L & \underline{likes} & L & \underline{kangaroos} & L \\
  NP & L & (S\setminus NP)/NP & S\setminus NP & NP & >A \\
  S & S & >A
\end{array}
\]
```
What’s parsing?

- Syntactic analysis of a string
- CFG, CCG, Dependency Grammar
What do we need for parsing?

A formalism

- CFG
- CCG
- Dependency Grammar
- RCG
- TAG
- LFG,
- HPSG
- ... [Müller, 2010]

A parsing algorithm

- Recursive Descent
- Earley
- CYK
- LR, Tomita
- ... [Shieber et al., 1995]
Topic 1: Viterbi parsing

Questions
Given a string and a CFG, . . .

► . . . how do we efficiently compute the best derivation?

► – or the best $n$ derivations? [Huang and Chiang, 2005]

► . . . how do we avoid accidentally discarding good derivations while pruning? [Klein and Manning, 2003]
Kim gives the kangaroos food
Kim gives the kangaroos food
Kim gives the kangaroos food
Kim gives the kangaroos food
Kim gives the kangaroos food
Kim gives the kangaroos food
Topic 2: Dependency parsing

Questions

▶ How do probabilistic models of DG work? [Eisner, 1996b]
▶ How do we deal with non-projective structures such as this one? [McDonald et al., 2005]
Topic 3: CCG parsing

What is CCG?

- Combinatory Categorial Grammar (Steedman)
- a fully lexicalized grammar formalism
- a few primitive categories (NP, S, Det) and infinitely many complex
- pure CG: only functional application
- CCG: combinatory rules: type-raising, functional composition

Advantages of CCG

- categories for non-standard constituents (e.g. non-constituent coordination)
- easy to do semantics: The semantic type of the logical form is entirely determined by the syntactic category
Topic 3: CCG parsing

Functional application only:

\[
\frac{Kim}{NP} \quad \frac{likes}{(S\setminus NP)/NP} \quad \frac{kangaroos}{NP} \\
\frac{(S\setminus NP)\setminus NP}{S\setminus NP} \quad \frac{S\setminus NP}{S} \quad \text{>A}
\]

With type-raising and functional composition:

\[
\frac{Kim}{NP} \quad \frac{likes}{(S\setminus NP)/NP} \quad \frac{kangaroos}{NP} \\
\frac{S/(S\setminus NP)}{S/np} \quad \frac{S/np}{S} \quad \text{>A}
\]

Daniel Quernheim

Parsing

April 17, 2012
Topic 3: CCG parsing

Questions

▶ How to avoid spurious ambiguity due to combinatory rules? [Eisner, 1996a]
▶ How to build a statistical CCG parser? [Clark and Curran, 2007]
Topic 4: TAG parsing

What is TAG?

- Tree Adjoining Grammar
- elementary trees are composed by substitution and adjunction
- can elegantly capture long-distance dependencies

```
NP | kangaroos
S  | NP | VP
  |   |  V
  |   | jump
VP | Adj | VP*
   | always
```
Question:

- How can TAG be parsed efficiently?  
  [Shieber et al., 1995, Schabes and Joshi, 1988]
- (For a variety of TAG:) [Kallmeyer and Satta, 2009]
Topic 5: RCG parsing

What is RCG?

- Range Concatenation Grammar [Boullier, 1999]
- even more powerful than TAG
- covers the entire PTIME class of languages
- rules make use of predicates over variables denoting ranges

Example RCG

\[
S(XYZ) \rightarrow A(X, Y, Z) \\
A(ax, ay, az) \rightarrow A(X, Y, Z) \\
A(bx, by, bz) \rightarrow A(X, Y, Z) \\
A(\varepsilon, \varepsilon, \varepsilon) \rightarrow \varepsilon
\]
Topic 5: RCG parsing

Questions

▶ How can we place sensible restrictions on this framework to stay tractable?
▶ How do we parse Earley style? [Kallmeyer et al., 2009]
▶ How are common syntactic phenomena modeled? [Boullier, 1999]
▶ Can RCGs be used for machine translation? [Søgaard, 2008]
References

Chinese numbers, MIX, scrambling, and Range Concatenation Grammars.

Wide-coverage efficient statistical parsing with CCG and log-linear models.

Efficient normal-form parsing for Combinatory Categorial Grammar.
In Proc. ACL, pages 79–86.

Three new probabilistic models for dependency parsing: An exploration.

Better k-best parsing.
In Proc. IWPT, pages 53–64.

An Earley parsing algorithm for Range Concatenation Grammars.

A polynomial-time parsing algorithm for TT-MCTAG.

In Proc. HLT-NAACL.

Non-projective dependency parsing using spanning tree algorithms.
In Proc. HLT/EMNLP.

Grammatiktheorie.
Stauffenburg.

An earley-type parsing algorithm for Tree Adjoining Grammars.
In Proc. ACL, pages 258–269.

Principles and implementation of deductive parsing.

Range concatenation grammars for translation.