Introduction to Natural Language Syntax and Parsing

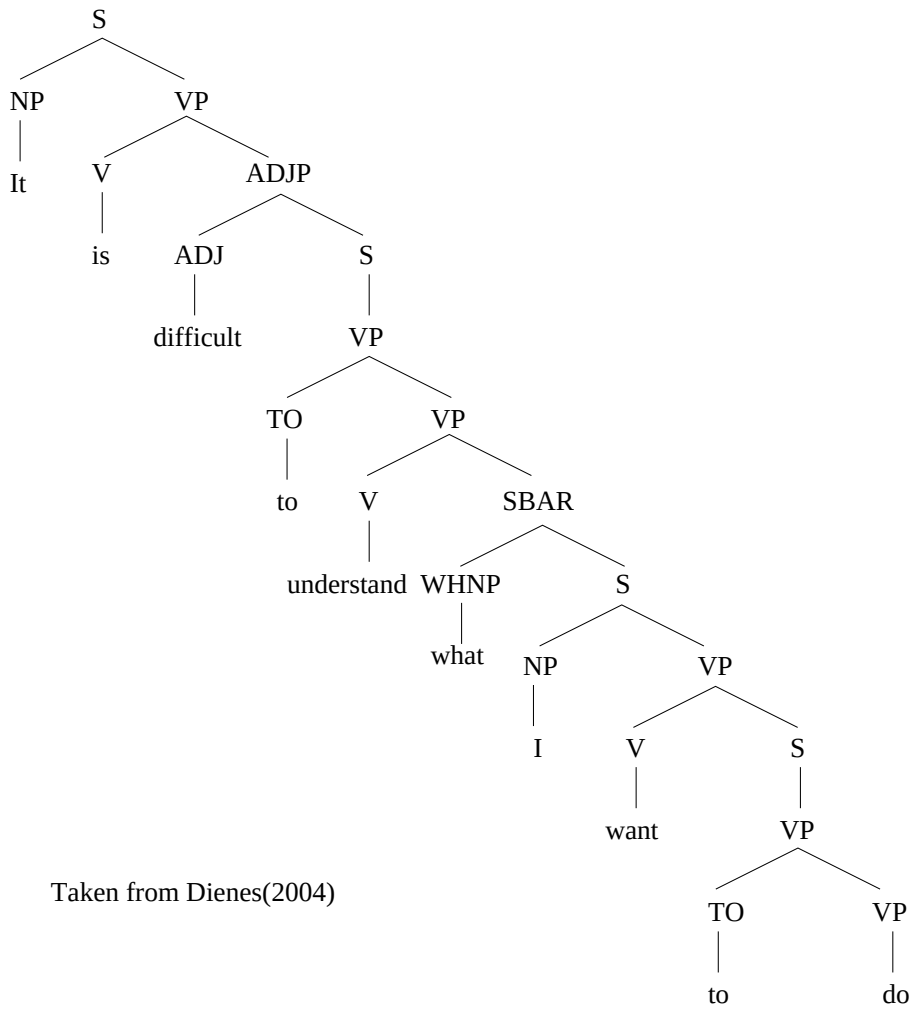
Lecture 4: Introduction to Parsing

Stephen Clark

Natural Language and Information Processing (NLIP) Group

sc609@cam.ac.uk

Parsing - Phrase Structure Trees



Taken from Dienes(2004)

Parsing - Dependency Trees

- More direct representation of how the words in a sentence are related, in terms of (labelled) edges between words
- Currently a popular form of parsing:
  - interesting algorithmic and learning problems;
  - useful for applications;
  - applicable to all languages (including eg free word order languages)
  - theory-neutral (to a large extent)

[picture on board]

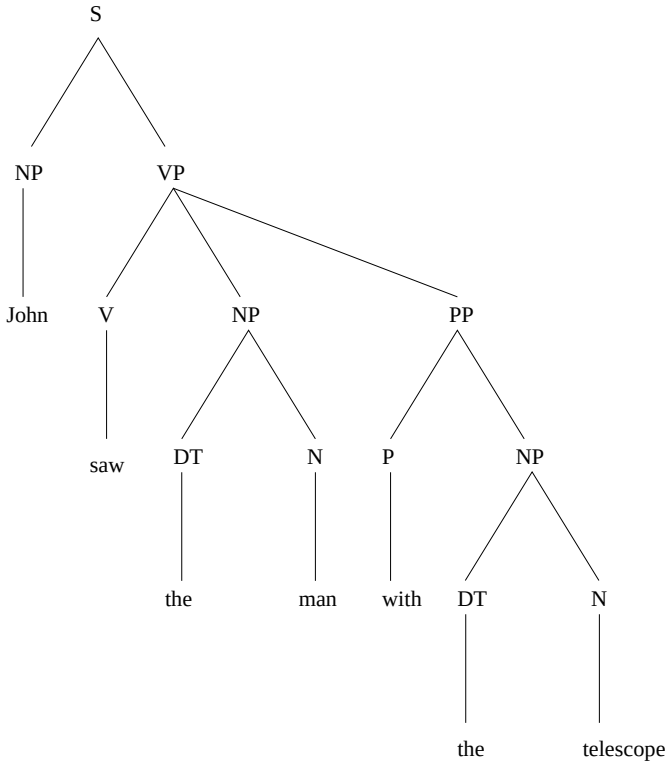
Characterisation of Parsing - 3 key questions

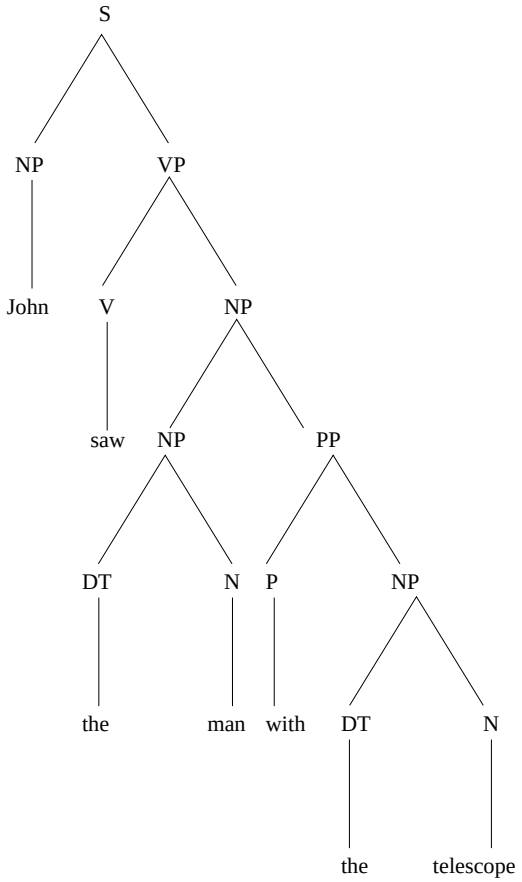
- What is the grammar of the natural language in question? Where does it come from?
- What is the algorithm which builds the possible parses for a sentence?
- What is the model for determining the plausibility of the parses (because there may be lots of alternatives)?

Why is Parsing Difficult?

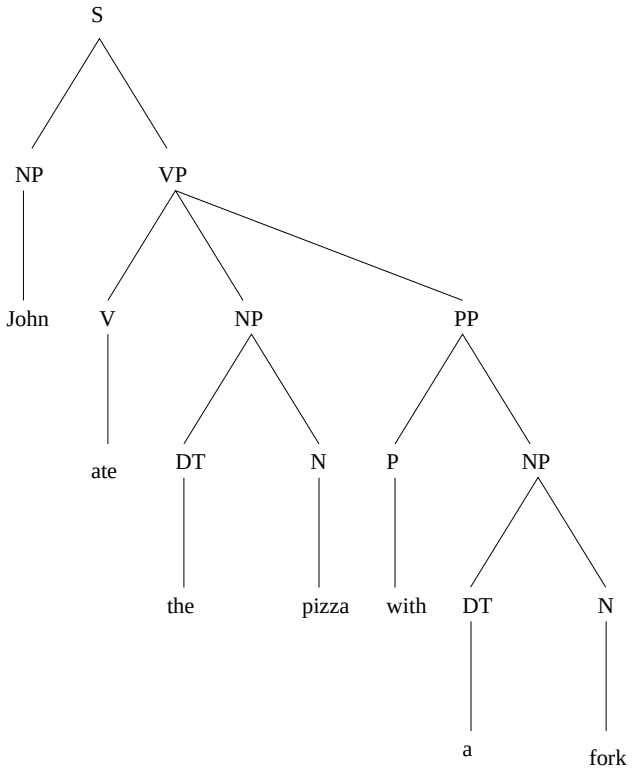
- Obtaining a *wide-coverage* grammar which can handle arbitrary real text is challenging
- Natural language is surprisingly **ambiguous**

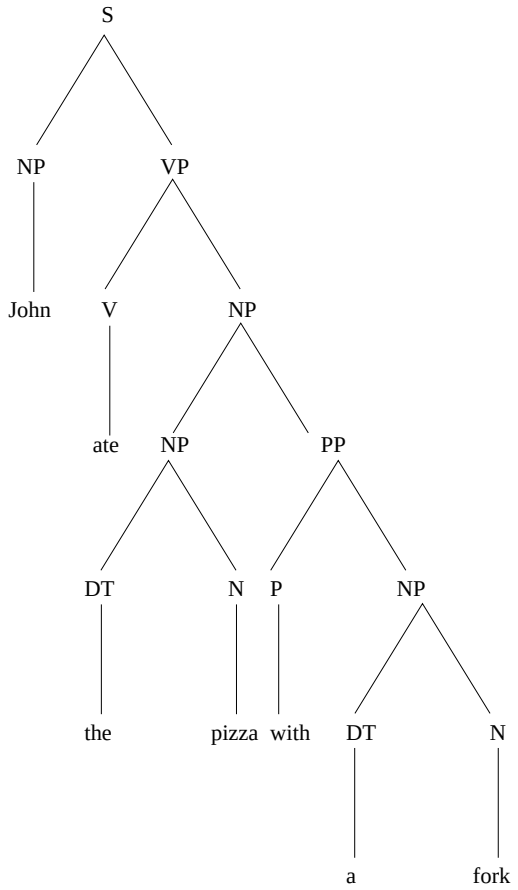
Syntactic Ambiguity



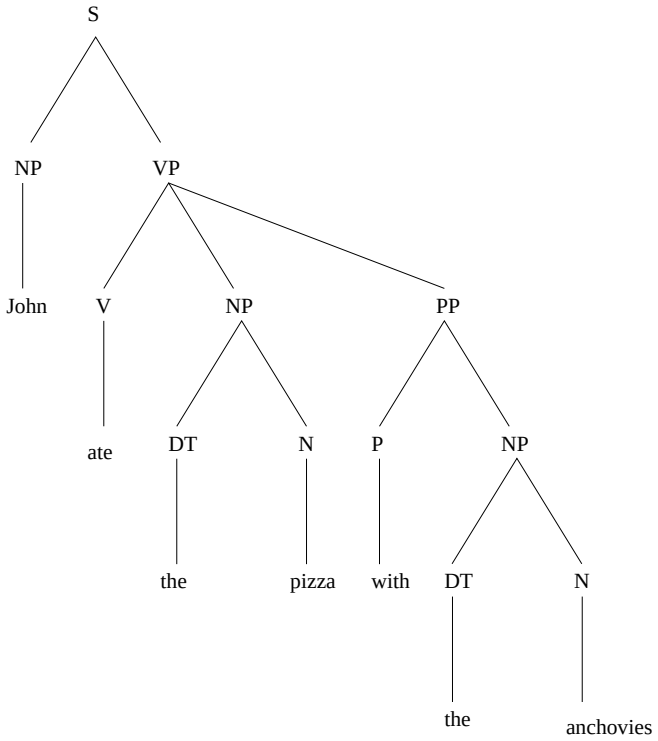


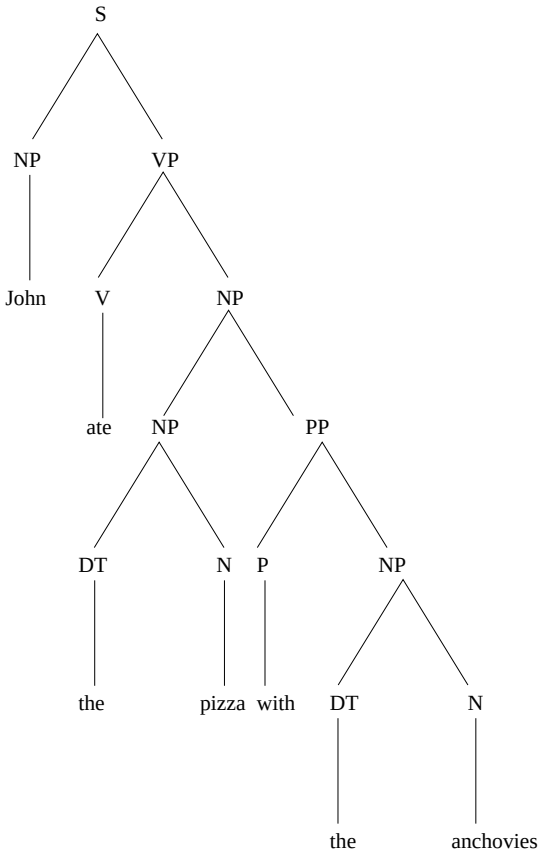
Syntactic Ambiguity: the problem is worse than you think





Syntactic Ambiguity: the problem is worse than you think





Syntactic Ambiguity: the problem is even worse than that

- Put the block in the box on the table 2 analyses
- Put the block in the box on the table beside the chair 5 analyses
- Put the block in the box on the table beside the chair before the table 14 analyses
- Put the block in the box on the table beside the chair before the table in the kitchen 42 analyses
- . . . 132 analyses
- . . . 469 analyses
- . . . 1430 analyses
- . . . 4862 analyses

Syntactic Ambiguity: the problem is even worse than that

- Previous sequence was the Catalan sequence; grows exponentially with the number of PPs
- Question: Ok, but we never see PPs stacked up like that in real sentences?
- Answer: but we do see other constructions with similar behaviour, eg coordination, and these various constructions stack up against each other

Syntactic Ambiguity: the problem is even worse than that

- Wider grammar coverage $\Rightarrow$ more analyses
- In practice this could mean millions (or more) of parses for a single sentence
  - difficult to imagine how productive these wide-coverage grammars can be without looking carefully at the output of a parser which uses one
- We need an efficient representation of this parse space
- And an efficient way to search it

Interesting Ambiguity Examples

- *The a are of I*
- *The cows are grazing in the meadow*
- *John saw Mary*

examples from Abney (1996)

Reading for Today's Lecture

- Chapters 9 and 10 of Manning and Schutze
- Steven Abney (1996), Statistical Methods and Linguistics, available from Abney's webpage