Parsing Components

Miriam Butt
October 2002

We'll go through the following basic parsing components:

- Phrase Structure Rules for a given set of sentences
- Possible Parsing Strategies/Algorithms
- Lexical Knowledge (Subcategorization Frames, POS, World Knowledge)
- Morphological Analysis
- Interaction of Morphology with Syntax

Phrase Structure Rules

Recall our first sample set for English.

S → NP VP

VP → V NP

NP → Det (ADJ) N

Plus some Lexical Items:

Mary N	green Adj	sees V	a Det
David N	big Adj	laughs V	the Det
dog N	cat N	gives V	heavy Det

Phrase Structure Rules

Phrase Structure Rules are *Regular Expressions*

Some Notation

( ) denotes an optional element

* The Kleene star allows for 0 to infinitely many elements

+ allows for 1 to infinitely many elements

Phrase Structure Rules

What sentences can this small grammar generate/parse?

What sentences can this small grammar not deal with?

How can we change the grammar to account for:

Mary laughs

Mary sees David.

Mary sees a big green cat.

Phrase Structure Rules

XLE contains a Toy English Grammar

This grammar does not allow for PP-Attachment ambiguity.

The girl saw the monkey with the telescope.

How can we change the grammar to give us two parses, rather than one?

XLE

The Ambiguity Problem Continued

Categorial Ambiguity

Flying planes can be dangerous.

V/Adj

Time flies like an arrow.	N V Adv Det N
Fruit flies like a banana.	Adj N V Det N

The Ambiguity Problem Continued

This kind of Ambiguity is difficult to deal with.

One helpful component is a good POS tagger/guesser.

Sometimes differing *Parsing Strategies* can also help.

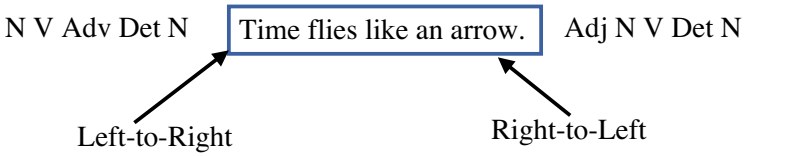
Parsing Strategies

Starting with a given string and a given grammar, a parser has several strategic options.

Left-to-Right vs. Right-to-Left

Bottom-Up vs. Top-Down

Parsing Strategies

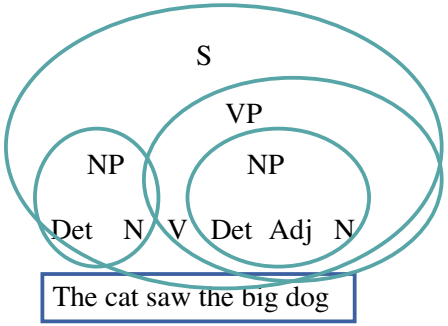


Start with the left edge of the string and work rightwards.

Start with the right edge of the string and work leftwards.

Parsing Strategies
Bottom Up

Start with the terminal elements, try to identify their POS and build them into constituents permitted by the grammar.

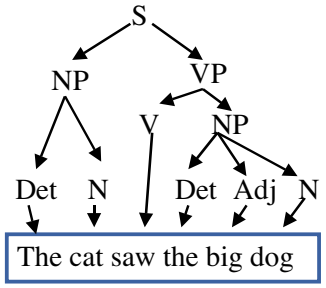


Parsing Strategies
Top down

Start with the top level phrase structure rule, expand it and try to fit the terminal elements with a possible expansion of the phrase structure rule.

1. NP → Det Adj N
2. NP → N
3. NP → Det N

1. VP → V
2. VP → V NP



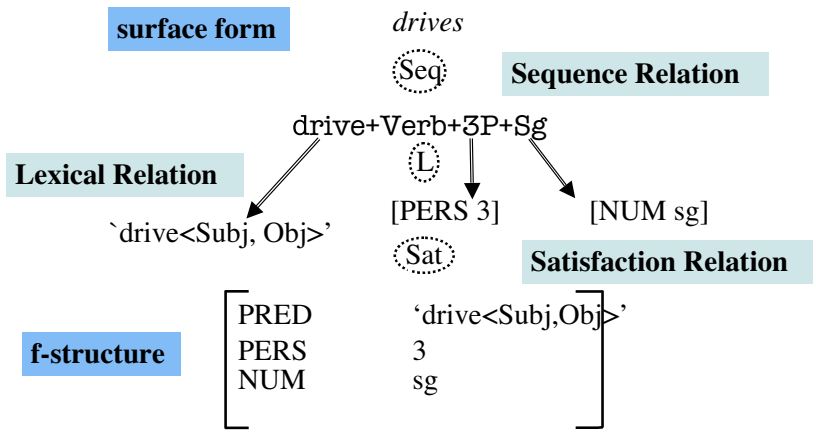
Lexicons

Typically Contain:

- Category Information (Terminal Node in Tree)
- Context Sensitive Featural Information
- Subcategorization Information
- Semantics (sometimes)

See CL Intro Slides (pp. 20-29)

XLE

FST Morphological Analyzers

The morphology-syntax interface for an LFG grammar: