

Parsing

CS395 GAI
Spring 2005

Overview

- What syntax does
- Grammars
- Feature structures
- Ambiguity
- Robustness

Syntax as serialization

- Natural language packs complex structural description into linear string of words
 - A first-order approximation
 - Ignores prosody, gestures, facial expressions...
 - Accurate enough for our purposes
- Parsing is part of the process of reconstructing the intended meaning
 - Often only the first step

Grammars

- Specify what combinations of structures are allowable in a language
- Consist of categories and rules
 - Categories typically syntactic categories
 - Exception: Semantic grammars
 - Rules describe legal combinations

A simple grammar

- $S \rightarrow NP VP$
- $VP \rightarrow V$
- $VP \rightarrow V NP$
- $VP \rightarrow ADJ V NP$
- $NP \rightarrow \textit{any proper name}$
- $NP \rightarrow DET NP$
- $NP \rightarrow DET ADJ NP$
- $DET \rightarrow \text{“a”} \mid \text{“the”}$

- N, V, ADJ, ADV – any items marked in the lexicon

A legal sentence

- “John served the hot dish.”

```
((noun John) (verb served) (det the)
      (adj hot) (noun dish))
```

```
((np (noun John)) (verb served)
      (np (det the) (adj hot) (noun dish)))
```

```
((np (noun John))
      (vp (verb served)
            (np (det the) (adj hot) (noun dish))))
```

```
(s (np (noun John))
      (vp (verb served)
            (np (det the) (adj hot) (noun dish))))
```

With the listener

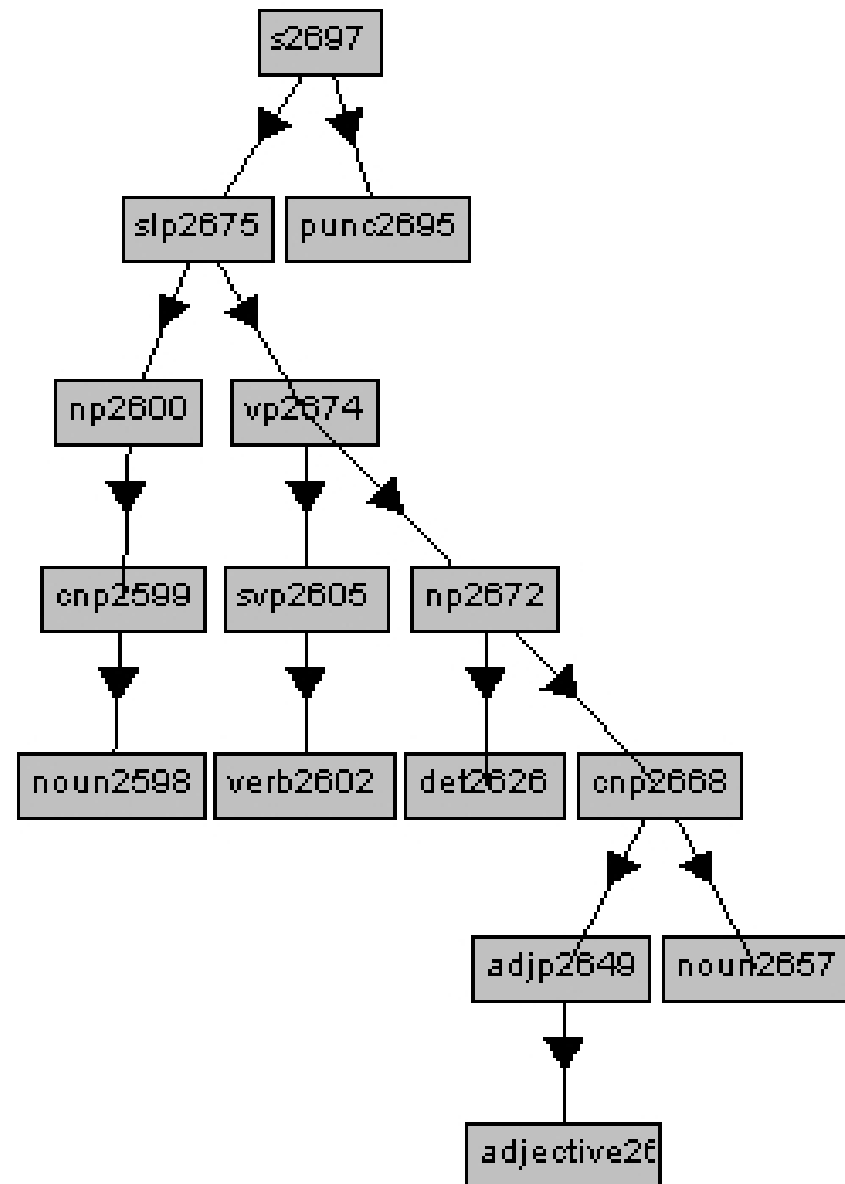
(ea::parse "John served the hot dish.")

1 parse result(s):

Complete sentence parses: 1

Sentence-level phrases: 0

Misc phrases & fragments: 0



Other sentences

- The food served John.
- John served.
- Serve John.

Complexity: Rules need to take into account properties of words

- Does the verb require a complement?
 - “John served.” versus “John put.”
- Is it singular or plural?
 - “Cats are stupid” versus “Cat is stupid”
- Is it a mass noun or a count noun?
 - “Water is” versus “Waters are”

Features

- Put additional constraints on what constituents can match with a rule
- Can also use to accumulate semantics

```
((slp (agr ?a) (var ?varvp)
      (:ACTION ?varvp) (:EVENT ?varvp) (:SUBJECT ?varnp)
      (sem (data::and ?semvp ?semnp))))
```

-slp->np-vp-

```
(np (agr ?a) (sem ?semnp) (var ?varnp))
(head (vp (inv -) (agr ?a) (var ?varvp) (sem ?semvp))))
```

Ambiguity

- Lexical ambiguity is one source
 - “hot dish” = warm food | casserole | attractive person
- Prepositional phrase attachment is another
 - “The man on the hill with the telescope saw the star”
 - “The hammer behind the door should go in the shelf inside the small room on the right.”

Exercise

- Divide up into pairs
 - One of you will be A, the other B
- Two scenarios will be presented
 - One of you will be the player, the other will be the NPC
 - There will be a specific goal that you are trying to jointly accomplish
 - All communication must be done in writing, taking turns

Scenario 1: Bomb Squad

- Bomb squad expert, typing over IM channel to a citizen who has found a bomb that will go off very shortly if not defused. Once the bomb is exposed to view (it is not visible to the citizen, a tip was received about it), there are five visible wires: red, green, blue, white, and black.
- A (Player):
 - Must communicate that the correct sequence of wires to cut is green, red, then blue.
 - Must keep the citizen focused on doing the right thing.
- B (NPC):
 - Must project fear/tension.
 - Make the player care about whether they live or die.
 - Throw enough roadblocks in their way that it takes about 10 conversational turns to defuse the bomb.

Scenario 2: Interstellar negotiations

- Contact has been made via hyperwave to an advanced civilization. The player is trying to negotiate a cure for cancer. The Wub is curious about this “aboriginal culture”
- A (NPC):
 - Find some commodity worth trading the cure for
 - Make the player work at closing the deal, but get it closed within 10 conversational turns
- B (Player):
 - Strike a deal to get the cure for cancer
 - Give away as little as you can get away with, since there are many other things that Earth would like to get from them, and there are only so many cultural artifacts that they will find interesting.

Properties of Grammars

- **Generality:** Range of sentences it analyzes correctly
- **Robustness:** Able to pull something out of partial understanding
 - Contrast with Selectivity: The range of non-sentences it identifies as problematic
- **Understandability:** How easily can someone understand the grammar?
 - Important for maintenance and extension

Robustness

- Many human utterances are not grammatical
 - “Teh shoos don’t fit.” (misspelling)
 - “the green one” (ellipses)
 - “Then he goes, like, ...” (dialect)
 - “It was a sketchy bar.” (language evolution)
- Need to pull as much out of the utterance as feasible

Typical approach

- When full parse fails, gather fragments
 - Bottom-up chart parsers dominate in part because they provide such partial analyses
 - Try local repairs (spell check, soundex)
- See how fragments fit into the current context
 - “Which wire should I cut?”
 - “Quick, make some noise to distract the guard!”

Moving toward semantics

- Parsing assembles the initial information needed for semantic interpretation
 - Entities introduced to stand for the things referred to. (`var` feature)
 - Semantic information associated with each word pulled out of the knowledge base
 - Grammatical roles used to fill in assertions
- Typically ambiguous at this stage
 - Multiple alternatives, due to different word senses
 - Further processing required to produce useful output