Syntactic Ambiguity

CSCI-GA.2590 – Lecture 3B

Ralph Grishman
Part-of-speech Ambiguity

• Many English words have 2 (or more) parts of speech
  • the last fast train speeds up station schedules

• produces lots of structures during parsing
Attachment Ambiguity

- PPs [prepositional phrases] can modify any of several prior constituents:

  I saw the man in the park

  ?
Grows rapidly

- Ambiguity grows rapidly with size of sentence

I saw the man in the park with a telescope

I saw the man on a hill in the park with a telescope
Grows rapidly

- Number of parses (and time) grows as n!

(n factorial ... faster than exponential)
Dynamic programming

• Group together all analyses of span \( <i, j> \) as grammar symbol X
  • factored representation of parse trees
• \( n^2/2 \) spans for an n-word sentence
• with binarized form of grammar, can parse in polynomial time (\( O(n^3) \))
CKY parse table

- one cell for each possible span
  - combine different derivations of same symbol with same span

```
[0, 1] [0, 2] [0, 3] [0, 4]
[1, 2] [1, 3] [1, 4]
[2, 3] [2, 4]
[3, 4]
```
Doesn’t solve the problem

• Dynamical programming allows us to generate multiple parses much faster, but doesn’t really solve the problem:
  – what do we do with all these parses?
  – how do we select the likely interpretation of the sentence?
Adding constraints

• our simple grammar generates lots of non-sentences (*Cat sleep.)

• can we reduce the number of parses by incorporating some constraints, such as:
  • number agreement
  • subcategorization
  • selection
Number Agreement

sentence → np-sing vp-sing | np-plur vp-plur
np-sing → ngroup-sing | ngroup-sing pp
ngroup-sing → n-sing | art-sing n-sing | art-sing adj n-sing
vp-sing → v-sing | v-sing np | v-sing vp | v-sing np pp
np-plur → ngroup-plur | ngroup-plur pp
ngroup-plur → n-plur | art-plur n-plur | art-plur adj n-plur
vp-plur → v-plur | v-plur np | v-plur vp | v-plur np pp
pp → p np

Cats sleep.  *Cats sleep.
Subcategorization

sentence $\rightarrow$ np vp

np $\rightarrow$ ngroup | ngroup pp

ngroup $\rightarrow$ n | art n | art adj n

vp $\rightarrow$ v-intrans | v-trans np | v-xcomp vp | v-ppcomp np pp

pp $\rightarrow$ p np

He sleeps.  *He sleeps dreams.
Little gain

• Each constraint enlarges grammar substantially
• Effect of multiple constraints is multiplicative
• Except for subcategorization, benefit is small – few bad parses are eliminated
What's left?

How do we really cope with ambiguity?

• Probabilistic grammar
  – to discuss in detail in a few weeks

• Partial parsing
  – to discuss next
  – only parse as much as we need
  – will need to resolve part-of-speech ambiguities first