

CMPT-413

Computational Linguistics

Anoop Sarkar
`http://www.cs.sfu.ca/~anoop`

January 13, 2011

Context Free Grammars and Ambiguity

<i>S</i>	→	<i>NP VP</i>
<i>VP</i>	→	<i>V NP</i>
<i>VP</i>	→	<i>VP PP</i>
<i>PP</i>	→	<i>P NP</i>
<i>NP</i>	→	<i>NP PP</i>
<i>NP</i>	→	<i>Calvin</i>
<i>NP</i>	→	<i>monsters</i>
<i>NP</i>	→	<i>school</i>
<i>V</i>	→	<i>imagined</i>
<i>P</i>	→	<i>in</i>

What is the analysis using the above grammar for:
Calvin imagined monsters in school

Context Free Grammars and Ambiguity

Calvin imagined monsters in school

```
(S (NP Calvin)
  (VP (V imagined)
    (NP (NP monsters)
      (PP (P in)
        (NP school))))))
```

```
(S (NP Calvin)
  (VP (VP (V imagined)
    (NP monsters))
    (PP (P in)
      (NP school))))
```

Which one is more plausible?

Ambiguity Kills (your parser)

natural language learning course

(run demos/parsing-ambiguity.py)

((natural language) (learning course))

((natural language) learning) course)

((natural (language learning)) course)

(natural (language (learning course)))

(natural ((language learning) course))

- ▶ Some difficult issues:
 - ▶ Which one is more plausible?
 - ▶ How many analyses for a given input?
 - ▶ Computational complexity of parsing language

Treebanks

- ▶ What is the CFG that can be extracted from this single tree:

```
(S  (NP (Det the) (NP man))
    (VP (VP (V played)
             (NP (Det a) (NP game)))
        (PP (P with)
             (NP (Det the) (NP dog))))))
```

PCFG

<i>S</i>	\rightarrow	<i>NP VP</i>	$c = 1$
<i>NP</i>	\rightarrow	<i>Det NP</i>	$c = 3$
<i>NP</i>	\rightarrow	<i>man</i>	$c = 1$
<i>NP</i>	\rightarrow	<i>game</i>	$c = 1$
<i>NP</i>	\rightarrow	<i>dog</i>	$c = 1$
<i>VP</i>	\rightarrow	<i>VP PP</i>	$c = 1$
<i>VP</i>	\rightarrow	<i>V NP</i>	$c = 1$
<i>PP</i>	\rightarrow	<i>P NP</i>	$c = 1$
<i>Det</i>	\rightarrow	<i>the</i>	$c = 2$
<i>Det</i>	\rightarrow	<i>a</i>	$c = 1$
<i>V</i>	\rightarrow	<i>played</i>	$c = 1$
<i>P</i>	\rightarrow	<i>with</i>	$c = 1$

- ▶ We can do this with multiple trees. Simply count occurrences of CFG rules over all the trees.
- ▶ A repository of such trees labelled by a human is called a TreeBank.

Ambiguity

- ▶ Part of Speech ambiguity

saw → noun

saw → verb

- ▶ Structural ambiguity: Prepositional Phrases

I saw (the man) with the telescope

I saw (the man with the telescope)

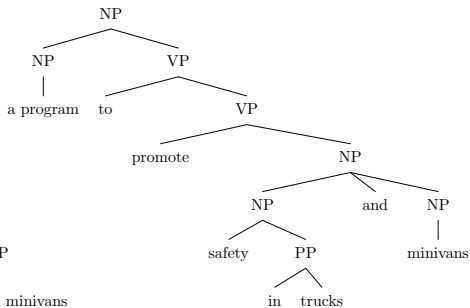
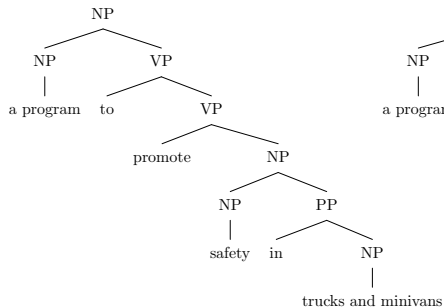
- ▶ Structural ambiguity: Coordination

a program to promote safety in ((trucks) and (minivans))

a program to promote ((safety in trucks) and (minivans))

((a program to promote safety in trucks) and (minivans))

Ambiguity ← attachment choice in alternative parses



Ambiguity in Prepositional Phrases

- ▶ noun attach: *I bought the shirt with pockets*
- ▶ verb attach: *I washed the shirt with soap*
- ▶ As in the case of other attachment decisions in parsing: it depends on the meaning of the entire sentence – needs world knowledge, etc.
- ▶ Maybe there is a simpler solution: we can attempt to solve it using heuristics or associations between words

Structure Based Ambiguity Resolution

- ▶ Right association: a constituent (NP or PP) tends to attach to another constituent immediately to its right (Kimball 1973)
- ▶ Minimal attachment: a constituent tends to attach to an existing non-terminal using the fewest additional syntactic nodes (Frazier 1978)
- ▶ These two principles make opposite predictions for prepositional phrase attachment
- ▶ Consider the grammar:

$$VP \rightarrow V NP PP \quad (1)$$

$$NP \rightarrow NP PP \quad (2)$$

for input: *I* [_{VP} *saw* [_{NP} *the man* . . . [_{PP} *with the telescope*],
RA predicts that the PP attaches to the NP, i.e. use rule (2),
and MA predicts V attachment, i.e. use rule (1)

Structure Based Ambiguity Resolution

- ▶ Garden-paths look structural:
The emergency crews hate most is domestic violence
- ▶ Neither MA or RA account for more than 55% of the cases in real text
- ▶ Psycholinguistic experiments using eyetracking show that humans resolve ambiguities as soon as possible in the left to right sequence using the words to disambiguate
- ▶ Garden-paths are caused by a combination of lexical and structural effects:
The flowers delivered for the patient arrived

Ambiguity Resolution: Prepositional Phrases in English

- Learning Prepositional Phrase Attachment: Annotated Data

v	n1	p	n2	Attachment
join	board	as	director	V
is	chairman	of	N.V.	N
using	crocidolite	in	filters	V
bring	attention	to	problem	V
is	asbestos	in	products	N
making	paper	for	filters	N
including	three	with	cancer	N
⋮	⋮	⋮	⋮	⋮

Prepositional Phrase Attachment

Method	Accuracy
Always noun attachment	59.0
Most likely for each preposition	72.2
Average Human (4 head words only)	88.2
Average Human (whole sentence)	93.2

Some other studies

- ▶ **Toutanova, Manning, and Ng, 2004:** 87.54% using some external knowledge (word classes)
- ▶ **Merlo, Crocker and Berthouzoz, 1997:** test on multiple PPs
 - ▶ generalize disambiguation of 1 PP to 2-3 PPs
 - ▶ 14 structures possible for 3PPs assuming a single verb
 - ▶ all 14 are attested in the Penn WSJ Treebank
 - ▶ 1PP: 84.3% 2PP: 69.6% 3PP: 43.6%
 - ▶ **This experiment is still only part of the real problem faced in parsing English**
 - ▶ Other sources of ambiguity in other languages