Lecture 6
CFGs – Good for your health or not?
Lord of the Loops:
Three loops to rule them all

Professor Robert C. Berwick
berwick@csail.mit.edu
Ambiguity is... everywhere

• Part of speech tag: dog/dog N/V
Syntactic Ambiguity & Human processing

- I told Bill that my dog died yesterday
- I told Bill that my dog will die yesterday
What *are* the possible syntactic relationships in a sentence?

What other relationships are needed?

6.863J/9.611J Fall 2012  Lecture 6
VP

X V' 

V

CAUSE

NP

recipient

GET

theme

VP

give

NP

John

GET

a book
Warlpiri: a ‘free word order language’ but…

Ngajulu-rlu ka-rna-rla punta-rni kurdu-ku karli
‘I am taking the boomerang from the child’
‘I tense-3-1 take the child-from boomerang’
A remarkable story about a remarkable person
A remarkable story about a remarkable person
Figure 3: (a) Processing hierarchy in Natural Grammar. Top: Schematic view of the different conditions. Bottom: Examples of stimulus items for each condition and schematic view of relation between subjects (S) and verbs (V) of (embedded) sentences. "Linear" stands for "no embedding". Dependent items are color-coded (red, green, blue).

(b) The linguistic description of a sentence used in the natural grammar study (Makuuchi et al. 2009). This sentence represents the most complex condition (Hierarchical Structure, long-distance dependency; compare Figure 2).

Key:
- ADV = adverb
- AUX = auxiliary
- C = clause
- COMP = complementizer
- INFL = inflection
- IP = inflectional phrase
- N = noun
- NP = noun phrase
- PAST = past tense
- REL = relative pronoun
- S = sentence
- V = verb
- VP = verb phrase

Syntactic hierarchy, as defined by the number of embeddings, activated Broca’s area in the inferior frontal gyrus (IFG). In addition, the left superior temporal gyrus (STG) and the superior temporal sulcus (STS) are also activated, indicating that these regions are part of the language network (Friederici et al. 2009). A region of interest analysis of the IFG (Makuuchi et al. 2009) revealed that the main effect of hierarchy was located in BA 44 as defined cytoarchitectonically according to Amunts et al. (1999). In contrast, working memory operationalized by the factor distance between the dependent elements activated the left inferior frontal sulcus located dorsally to Broca’s area (see Table 1 and Figure 4). A functional connectivity analysis revealed that these two areas strongly interact during processing multiple embedded sentences.

Figure 4: Schematic view of activation pattern for the main effect of hierarchy in the language domain. For Artificial Grammar I and II, the main effect of hierarchy was found in Broca’s area (BA 44/45) (Friederici et al. 2006, Bahlmann et al. 2008). For the natural grammar, the main effect of hierarchy was located in BA 44 (Makuuchi et al. 2009) and in the posterior superior temporal gyrus (pSTG) extending into the superior temporal sulcus (Friederici et al. 2009).

Key: BA = Brodman Area; CS central sulcus; IFS = inferior frontal sulcus; STG = superior temporal gyrus.
Activation areas are different for ‘normal’ language as opposed to ‘puzzle solving’

Maria comprè le no caffè

Maria comprè le caffè
What about engineered parsers?
A CFG challenge

JOHN GAVE MARY AND BILL TO FRED BOOKS THAT LOOKED REMARKABLY SIMILAR
JOHN HUMMED AND MARY SANG AT EQUAL VOLUMES
JOHN HUMMED AND MARY SANG THE SAME TUNE
I HAVE BEEN WONDERING WHETHER BUT WOULD NOT WANT TO STATE THAT YOUR THEORY IS CORRECT
I CAN TELL YOU WHEN BUT I CAN NOT TELL YOU WHY BILL LEFT ME
I LIKE BUT BILL DOES NOT LIKE TO VISIT NEW CITIES
JOHN HAS CLAIMED BUT I DO NOT BELIEVE THAT BILL IS A COMMUNIST
MARY USED TO BE AND BILL STILL IS VERY SUSPICIOUS
MARY SAID SHE WOULD AND BILL ACTUALLY DID EAT A RAW EGGPLANT
BILL MAY BE AND JOHN CERTAINLY IS A WEREWOLF
BILL OFFERED AND MARY GAVE JOHN A CADILLAC
BILL CAUGHT AND MARY KILLED THE RABID DOG
THE WOMAN WHO WAS HERE BELIEVED THAT THE MAN WAS ILL
THE WOMAN BELIEVED THAT THE GUY WHO WAS HERE WAS ILL
THE WOMAN BELIEVED THAT THE GUY WAS ILL WHO WAS HERE
TO WHICH CITY AND WHICH CONFERENCE DID BILL GO TO
WHICH CITY AND TO WHICH CONFERENCE DID BILL GO TO
WHICH CITY AND WHICH CONFERENCE DID BILL GO TO TO
TO WHICH CITY AND WHICH CONFERENCE DID BILL GO
TO WHICH CITY AND TO WHICH CONFERENCE DID BILL GO
ON WHICH TABLE AND UNDER WHICH FLOWER POT DID MARY PUT THE KEYS
WHERE AND WHEN DID BILL PUT THE BOOK
WHICH BOOK AND WHICH PENCIL DID MARY BUY
MARY ASKED WHO AND WHAT BOUGHT
MARY ASKED WHO AND WHERE BILL HAD SEEN
I WONDER WHO MARY LIKES AND HOPES WILL WIN
I WONDER WHO BILL SAW AND LIKED MARY
I WONDER WHO SAW BILL AND LIKED MARY
I WONDER WHO BILL SAW AND MARY LIKED
John gave Mary, Bill to Fred that books that looked remarkably similar.
Which violins are too difficult to play on these sonatas?
Phrase structure rules can be eliminated!

Let’s do this one step at a time.
First: Why do we have the first rule but not the second?

\[
\text{NP} \rightarrow \ldots \ 	ext{NN} \ldots
\]

\[
\text{NP} \rightarrow \ldots \ 	ext{V} \ldots
\]

Both are equally ‘costly’ in terms of ‘minimum description length’

But the second one is impossible, it seems
1970

- Whenever a phrase has an XP, it has an X (1970)
- Whenever a phrase has an X, it has an XP (1982)
- (together, a bi-conditional)
- There are no phrases, just lexical items, with their properties (which properties?) projected
The short shelf life of context-free grammars

- Introduced in 1963 to generate the ‘deep structure’ of a language (i.e., predicate-argument relations)
- But by 1965, they half-disappear...and by 1970...they’re nearly history...by 1989, they are gone!
- What happened and why?
What happened?

CFG has rules like this:

\[ VP \rightarrow V_1 \ NP \ (\text{eat ice-cream}) \]
\[ VP \rightarrow V_3 \ PP (\text{talk about}) \]
\[ VP \rightarrow V_4 \ S \ (\text{think that John likes me}) \]

...

Dictionary (lexicon) has rules English learner has internalized like this:

\[
\begin{align*}
\text{see: } & ___ \ NP \\
\text{think: } & ___ \ S \\
\text{solve: } & ___ \ NP \\
\text{talk: } & ___ \ PP \\
\end{align*}
\]

but crucially individual verbs include other idiosyncratic information! (eg, can solve a problem, but not solve a hotdog....)
What’s the point?

• The existence of rules like:
  VP → V NP
  VP → V PP
  VP → V S

*Implies* the existence of a lexical entry like *see*: ___NP

Why? Because if there were no entry like this, the rule would never have a chance to operate. (If no verb took a direct object, how would speakers ever know?)

But this means the rules & lexicon are redundant

If we have to choose between getting rid of the rules, and getting rid of the lexicon, we must get rid of the rules, because we don’t want to get rid of idiosyncratic facts. Deep point: CFG rules are entirely superfluous
But let’s work with CFGs anyway…

• How can we *parse* sentences efficiently with CFGs?
• Top-down or bottom-up?
• How to avoid the bottomless recursive pit?
• How to avoid computing the same thing more than once?
• How to integrate top-down and bottom-up information sources?
Learn to share!

\[
NP \rightarrow NP\ NP
\]
\[
NP \rightarrow \text{natural} | \text{language} | \text{processing} | \text{book}
\]
A toy grammar example

<table>
<thead>
<tr>
<th>S</th>
<th>→</th>
<th>NP</th>
<th>VP</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>→</td>
<td>NP</td>
<td>VB</td>
</tr>
<tr>
<td>VP</td>
<td>→</td>
<td>$V_i$</td>
<td></td>
</tr>
<tr>
<td>VP</td>
<td>→</td>
<td>$V_t$</td>
<td>NP</td>
</tr>
<tr>
<td>NP</td>
<td>→</td>
<td>DT</td>
<td>NN</td>
</tr>
<tr>
<td>NP</td>
<td>→</td>
<td>NP</td>
<td>PP</td>
</tr>
<tr>
<td>PP</td>
<td>→</td>
<td>P</td>
<td>NP</td>
</tr>
<tr>
<td>$V_i$</td>
<td>→</td>
<td>sleeps</td>
<td></td>
</tr>
<tr>
<td>$V_t$</td>
<td>→</td>
<td>saw</td>
<td></td>
</tr>
<tr>
<td>NN</td>
<td>→</td>
<td>guy</td>
<td></td>
</tr>
<tr>
<td>NN</td>
<td>→</td>
<td>dog</td>
<td></td>
</tr>
<tr>
<td>NN</td>
<td>→</td>
<td>person</td>
<td></td>
</tr>
<tr>
<td>NN</td>
<td>→</td>
<td>telescope</td>
<td></td>
</tr>
<tr>
<td>DT</td>
<td>→</td>
<td>the</td>
<td></td>
</tr>
<tr>
<td>IN</td>
<td>→</td>
<td>with</td>
<td></td>
</tr>
<tr>
<td>IN</td>
<td>→</td>
<td>in</td>
<td></td>
</tr>
</tbody>
</table>

0 the 1 dog 2 saw 3 the 4 man 5
Lord of the Loops:
Order of operations – CKY algorithm

Loop 1: Go *Across* columns, word by word, index $j$

Loop 2: Go *Up a* column, row by row, index $i$
Lord of the Loops:
Order of operations – CKY algorithm

Loop 1: Go Across columns, word by word, index $j$

Loop 2: Go Up a column, row by row, index $i$

Loop 3: Paste 2 previous entries (subtrees) together, index $k$
CKY: diagonal elements = the words

the
dog
saw
the
man
function CKY-PARSE(words, grammar) returns table

for $j \leftarrow$ from 1 to LENGTH(words) do
  $table[j - 1, j] \leftarrow \{ A \mid A \rightarrow \text{words}[j] \in \text{grammar} \}$
  for $i \leftarrow$ from $j - 2$ downto 0 do
    for $k \leftarrow i + 1$ to $j - 1$ do
      $table[i, j] \leftarrow table[i, j] \cup$
      $\{ A \mid A \rightarrow BC \in \text{grammar},$
      $B \in table[i, k],$
      $C \in table[k, j] \}$
CKY: diagonal elements then filled as POS tags
How many entries?