CMPT-882: Statistical Learning of Natural Language

Lecture #12

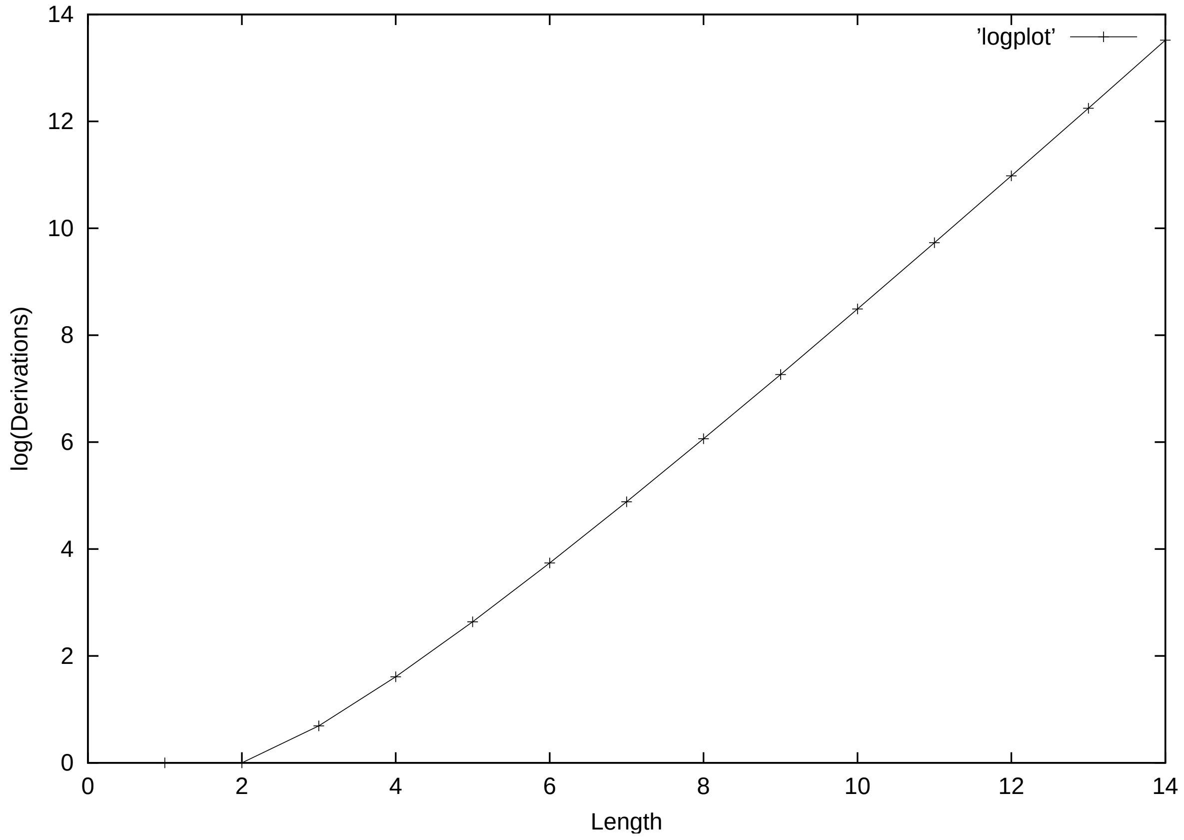
Anoop Sarkar
anoop@cs.sfu.ca
http://www.sfu.ca/~anoop

- Coping with syntactic ambiguity or how to put the block in the box on the table. (1982). Kenneth Church and Ramesh Patil. Computational Linguistics 8:139-49.

- Prepositional Phrase Attachment through a Backed-Off Model (1995). Michael Collins and James Brooks. Proceedings of the Third Workshop on Very Large Corpora WVLC-95.

Number of derivations grows exponentially

e.g. `L(G) = a a ...` for `G = S` $\longrightarrow$ `S S`



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 – the so-called AI complete problem
- First we give a precise characterization of the problem and then we try to solve it using statistical associations between words

Syntactic Ambiguity: (Church and Patil 1982)

- Algebraic character of parse derivations

- Power Series for grammar for coordination (more general than PPs):
  `NP` $\longrightarrow$ `cabbages` $|$ `kings` $|$ `NP and NP`

```
NP = cabbages + cabbages and kings
     + 2 (cabbages and cabbages and kings)
     + 5 (cabbages and kings and cabbages and kings)
     + 14 ...
```

Syntactic Ambiguity: (Church and Patil 1982)

- Coefficients equal the number of parses for each NP string
- These ambiguity coefficients are Catalan numbers:

$$Cat(n) = \binom{2n}{n} - \binom{2n}{n-1}$$

- $\binom{a}{b}$ is the *binomial coefficient*

$$\binom{a}{b} = \frac{a!}{(b!(a-b)!)}$$

Syntactic Ambiguity: (Church and Patil 1982)

- Why Catalan numbers? Cat(n) is the number of ways to parenthesize an expression of length n with two conditions:

  1. there must be equal numbers of open and close parens

  2. they must be properly nested so that an open precedes a close

- So the first term counts $2n$ parens with equal number of open and close, while the second term subtracts those that are not properly nested:

$$Cat(n) = \binom{2n}{n} - \binom{2n}{n-1}$$

Syntactic Ambiguity: (Church and Patil 1982)

- $Cat(n)$ also provides exactly the number of parses for the sentence:

```
John saw the man
         on the hill
         with the telescope
```

with 9 PPs: Cat(9) $= 4862$ parse trees

- Other sub-grammars are simpler:

$$
\begin{aligned}
ADJP &\rightarrow adj\ ADJP \ \mid \ \epsilon \\
ADJP &= 1 + adj + adj^2 + adj^3 + \ldots \\
ADJP &= \tfrac{1}{1-adj}
\end{aligned}
$$

Syntactic Ambiguity: (Church and Patil 1982)

- Now consider power series of combinations of sub-grammars:
  `S = NP · VP`

  ```
  ( The number of products over sales ... )
  ( is near the number of sales ... )
  ```

- Both the `NP` subgrammar and the `VP` subgrammar power series have Catalan coefficients

Syntactic Ambiguity: (Church and Patil 1982)

- The power series for the `S` $\rightarrow$ `NP VP` grammar is the multiplication:

$$(\; N \; \sum_i Cat_i \; (\; P \; N \;)^i) \; \cdot \; (\; is \; \sum_j Cat_j (\; P \; N \;)^j)$$

- In a parser for this grammar, this leads to a cross-product:

$$L \; \times \; R = \{(\; l, \; r \;) \mid l \in L \; \& \; r \in R \; \}$$

Syntactic Ambiguity: (Church and Patil 1982)

- A simple change:

```
Is ( The number of products over sales ... )
   ( near the number of sales ... )
```

$$= \texttt{Is}\ N\ \sum_i Cat_i\,(\ P\ N\ )^i)\ \cdot\ (\sum_j Cat_j(\ P\ N\ )^j)$$
$$= \texttt{Is}\ N\ \sum_i \sum_j Cat_i\, Cat_j(\ P\ N\ )^{i+j}$$
$$= \texttt{Is}\ N\ \sum_{i+j} Cat_{i+j+1}\,(\ P\ N\ )^{i+j}$$

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:
  e.g. in $\mathrm{I}\ [_{VP}\ \mathrm{saw}\ [_{NP}\ \mathrm{the\ man} \ldots [_{PP}\ \mathrm{with\ the\ telescope}\ ]$,
  RA predicts that the PP attaches to the NP,
  and MA predicts VP attachment

Structure Based Ambiguity Resolution

- Garden-paths look structural:
  The horse raced past the barn fell

- 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 lexical and not structural:
  The flowers delivered for the patient arrived

Ambiguity Resolution: Prepositional Phrases in English

- Statistical Methods for 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

If $p(1 \mid v, n1, p, n2) >= 0.5$ choose noun attachment

$$
\begin{aligned}
p(1 \mid v, n1, p, n2) \;=\;& \lambda(c_1) && \cdot\; p(1 \mid c_1 = v, n1, p, n2) \\
+\;& \lambda(c_2 + c_3 + c_4) && \cdot\; p(1 \mid c_2 = v, n1, p) \\
& && \cdot\; p(1 \mid c_3 = v, p, n2) \\
& && \cdot\; p(1 \mid c_4 = n1, p, n2) \\
+\;& \lambda(c_5 + c_6 + c_7) && \cdot\; p(1 \mid c_5 = v, p) \\
& && \cdot\; p(1 \mid c_6 = n1, p) \\
& && \cdot\; p(1 \mid c_7 = p, n2) \\
+\;& \lambda(c_8) && \cdot\; p(1 \mid c_8 = p) \\
+\;& (1 - \textstyle\sum_i \lambda(c_i)) && \cdot\; 1.0 \;\; \text{(default is noun attachment)}
\end{aligned}
$$

Katz Back-off Smoothing

1. If $f(v, n1, p, n2) > 0$ and $\widehat{p} \neq 0.5$

$$\widehat{p}(1 \mid v, n1, p, n2) = \frac{f(1, v, n1, p, n2)}{f(v, n1, p, n2)}$$

2. Else if $f(v, n1, p) + f(v, p, n2) + f(n1, p, n2) > 0$ and $\widehat{p} \neq 0.5$

$$\widehat{p}(1 \mid v, n1, p, n2) = \frac{f(1, v, n1, p) + f(1, v, p, n2) + f(1, n1, p, n2)}{f(v, n1, p) + f(v, p, n2) + f(n1, p, n2)}$$

3. Else if $f(v,p) + f(n1,p) + f(p,n2) > 0$

$$\widehat{p}(1 \mid v, n1, p, n2) = \frac{f(1,v,p) + f(1,n1,p) + f(1,p,n2)}{f(v,p) + f(n1,p) + f(p,n2)}$$

4. Else if $f(p) > 0$

$$\widehat{p}(1 \mid v, n1, p, n2) = \frac{f(1,p)}{f(p)}$$

5. Else $\widehat{p}(1 \mid v, n1, p, n2) = 1.0$

Prepositional Phrase Attachment: (Collins and Brooks 1995)

- Lexicalization helps disambiguation by capturing selectional preferences
  (Black et al. 1994; Magerman 1995)

- Smoothing deals with sparse data and improves prediction
  we ignore word senses here; cf. (Stetina and Nagao 1998)

- Uses the head of the phrase (e.g. prep) as privileged

- Similar insights led to lexicalization of grammars in mathematical linguistics and all-paths parsing; cf. TAG, CCG

Prepositional Phrase Attachment: (Collins and Brooks 1995)

- **Results**: 84.5% accuracy
  with the use of some limited word classes for dates, numbers, etc.

- Adding word sense disambiguation increases accuracy to 88%
  (Stetina and Nagao 1998)

- Can we improve on parsing performance using Probabilistic CFGs by using the insights detailed above

Two other studies

- **Brill and Resnik 1994**:
  use transformation based learning for PP attachment
  80.8% with words; with Wordnet classes: 81.8%
  only 266 transformations learned
  automatically learned importance of preposition (assumed in CB95)

- **Merlo, Crocker and Berthouzoz 1997**:
  test on multiple PPs, generalize the 2 PP case
  14 structures possible for 3PPs assuming a single verb: all 14 are attested in the Treebank
  same model as CB95; but generalized to dealing with upto 3PPs
  1PP: 84.3%   2PP: 69.6%   3PP: 43.6%
  this is still not the real problem faced in parsing natural language