Two views of linguistic structure: 1. Constituency (phrase structure)

- How do we know what is a constituent? (Not that linguists don't argue about some cases.)
  - Substitution/expansion/pro-forms:
    - *I sat on the box/right on top of the box/there*.
  - Distribution: a constituent behaves as a unit that can appear in different places:
    - *John talked to the children* [about drugs].
    - *John talked [about drugs] to the children*.
    - *John talked drugs to the children about*.
  - Coordination, regular internal structure, no intrusion, fragments, semantics,...
  - John [drove to the store] and [bought a bike].

Two views of linguistic structure: 2. Dependency structure

- Dependency structure shows which words depend on (modify or are arguments of) which other words.

The boy put the tortoise on the rug
Statistical Natural Language Parsing

Parsing: ambiguity and the rise of data and statistics

Pre 1990 ("Classical") NLP Parsing

- Wrote symbolic grammar (CFG or often richer) and lexicon
  \[ S \rightarrow NP \quad NP \rightarrow \text{interest} \]
  \[ NP \rightarrow (DT) \quad NNS \rightarrow \text{rates} \]
  \[ NP \rightarrow NNS \quad NNS \rightarrow \text{rates} \]
  \[ NP \rightarrow NNP \quad VBP \rightarrow \text{interest} \]

- Used grammar/proof systems to prove parses from words

  This scaled very badly and didn’t give coverage. For sentence:
  \[ \text{Fed raises interest rates 0.5\% in effort to control inflation} \]
  - Minimal grammar: 36 parses
  - Simple 10 rule grammar: 592 parses
  - Real-size broad-coverage grammar: millions of parses

Ambiguity: PP attachments

The boy ate the dessert with a spoon/cherry

Attachment ambiguities

- A key parsing decision is how we ‘attach’ various constituents
  - PPs, adverbial or participial phrases, infinitives, coordinations, etc.

  The board approved [its acquisition] [by Royal Trustco Ltd.]

  [of Toronto]

  [for $27 a share]

  [at its monthly meeting].

- Catalan numbers: \[ C_n = \frac{(2n)!}{(n+1)!n!} \]
- An exponentially growing series, which arises in many tree-like contexts:
  - E.g., the number of possible triangulations of a polygon with \( n + 2 \) sides
  - Turns up in triangulation of probabilistic graphical models….

Classical NLP Parsing: The problem and its solution

- Categorical constraints can be added to grammars to limit unlikely/weird parses for sentences
  - But the attempt makes the grammars not robust
    - In traditional systems, commonly 30% of sentences in even an edited text would have no parse.
- A less constrained grammar can parse more sentences
  - But simple sentences end up with ever more parses with no way to choose between them
- We need mechanisms that allow us to find the most likely parse(s) for a sentence
  - Statistical parsing lets us work with very loose grammars that admit millions of parses for sentences but still quickly find the best parse(s)

The rise of annotated data: The Penn Treebank

[Marcus et al. 1993, Computational Linguistics]
The rise of annotated data

- Starting off, building a treebank seems a lot slower and less useful than building a grammar
- But a treebank gives us many things
  - Reusability of the labor
  - Many parsers, POS taggers, etc.
  - Valuable resource for linguistics
  - Broad coverage
  - Frequencies and distributional information
  - A way to evaluate systems

Top-down parsing

Two problems to solve:
1. Repeated work...

Shift-reduce parsing: one path

Two problems to solve:
1. Repeated work...

What other search paths are there for parsing this sentence?
Two problems to solve:

1. Choosing the correct parse
   - How do we work out the correct attachment:
     - She saw the man with a telescope
     - Moscow sent more than 100,000 soldiers into Afghanistan ...
     - Our statistical parsers will try to exploit such statistics.
   - Is the problem ‘AI complete’? Yes, but ...
   - Words are good predictors of attachment
     - Even absent full understanding
   - Sydney Water breached an agreement with NSW Health ...

2. Choosing the correct attachment:
   - E.g.,
     - She saw the man with a telescope
     - Moscow sent more than 100,000 soldiers into Afghanistan ...
     - Our statistical parsers will try to exploit such statistics.

A simple prediction

- Use a likelihood ratio:
  - E.g.,
    - \[ \text{LR}(v, a, p) = \frac{P(p|v)}{P(p|a)} \]
  - \[ P(\text{with} | \text{agreement}) = 0.15 \]
  - \[ P(\text{with} | \text{breach}) = 0.02 \]
  - \[ \text{LR}(\text{breach, agreement, with}) = 0.13 \]
  - Choose noun attachment

A problematic example

- Chrysler confirmed that it would end its troubled venture with Maserati.
- Should be a noun attachment but get wrong answer:
  - \[ P(\text{with} | v) = \frac{607}{5156} \approx 0.118 > \frac{155}{1442} \approx 0.107 \]

A problematic example

- What might be wrong here?
  - If you see a V NP PP sequence, then for the PP to attach to the V, then it must also be the case that the NP doesn’t have a PP (or other postmodifier)
  - Since, except in extrapolation cases, such dependencies can’t cross
  - Also, the verb must take an NP object
  - ‘Unlike cases like ‘end with a bang’
  - Parsing allows us to factor in and integrate such constraints.

Human parsing

- Humans often do ambiguity maintenance
  - Have the police … eaten their supper?
  - come in and look around.
  - taken out and shot.
- But humans also commit early and are “garden pathed”:
  - The man who hunts ducks out on weekends.
  - The cotton shirts are made from grows in Mississippi.
A phrase structure grammar

\[ S \rightarrow NP \ VP \quad N \rightarrow \text{people} \]
\[ VP \rightarrow V \ NP \quad N \rightarrow \text{fish} \]
\[ VP \rightarrow V \ NP \ PP \quad N \rightarrow \text{tanks} \]
\[ NP \rightarrow NP \ PP \quad V \rightarrow \text{people} \]
\[ NP \rightarrow N \quad V \rightarrow \text{fish} \]
\[ NP \rightarrow e \quad V \rightarrow \text{tanks} \]
\[ PP \rightarrow P \ NP \quad P \rightarrow \text{with} \]

\begin{align*}
\text{people} & \quad \text{fish} \\
\text{fish} & \quad \text{tanks} \\
\text{people} & \quad \text{fish} \quad \text{with} \quad \text{rods}
\end{align*}

A phrase structure grammar

\[ S \rightarrow NP \ VP \quad N \rightarrow \text{people} \]
\[ VP \rightarrow V \ NP \quad N \rightarrow \text{fish} \]
\[ VP \rightarrow V \ NP \ PP \quad N \rightarrow \text{tanks} \]
\[ NP \rightarrow NP \ PP \quad V \rightarrow \text{people} \]
\[ NP \rightarrow N \quad V \rightarrow \text{fish} \]
\[ NP \rightarrow e \quad V \rightarrow \text{tanks} \]
\[ PP \rightarrow P \ NP \quad P \rightarrow \text{with} \]

\begin{align*}
\text{people} & \quad \text{fish} \\
\text{fish} & \quad \text{tanks} \\
\text{people} & \quad \text{fish} \quad \text{with} \quad \text{rods}
\end{align*}

Phrase structure grammars in NLP

\[ G = (T, C, N, S, R, P) \]
\[ T \text{ is a set of terminal symbols} \]
\[ C \text{ is a set of preterminal symbols} \]
\[ N \text{ is a set of nonterminal symbols} \]
\[ S \text{ is the start symbol } (S \in N) \]
\[ L \text{ is the lexicon, a set of items of the form } X \rightarrow x \]
\[ X \in P \text{ and } x \in T \]
\[ R \text{ is the grammar, a set of items of the form } X \rightarrow \gamma \]
\[ X \in N \text{ and } \gamma \in (N \cup C)^* \]
\[ \text{By usual convention, } S \text{ is the start symbol, but in statistical NLP, we usually have an extra node at the top (ROOT, TOP)} \]
\[ \text{We usually write } e/e \text{ for an empty sequence, rather than nothing} \]

A PCFG

\[ S \rightarrow NP \ VP \quad 1.0 \quad N \rightarrow \text{people} \quad 0.5 \]
\[ VP \rightarrow V \ NP \quad 0.6 \quad N \rightarrow \text{fish} \quad 0.2 \]
\[ VP \rightarrow V \ NP \ PP \quad 0.4 \quad N \rightarrow \text{tanks} \quad 0.2 \]
\[ NP \rightarrow NP \ NP \quad 0.1 \quad N \rightarrow \text{rods} \quad 0.1 \]
\[ NP \rightarrow NP \ PP \quad 0.2 \quad V \rightarrow \text{people} \quad 0.1 \]
\[ NP \rightarrow N \quad 0.7 \quad V \rightarrow \text{fish} \quad 0.6 \]
\[ PP \rightarrow P \ NP \quad 1.0 \quad V \rightarrow \text{tanks} \quad 0.3 \]
\[ P \rightarrow \text{with} \quad 1.0 \]

[With empty NP removed so less ambiguous]
The probability of trees and strings

- $P(t)$ – The probability of a tree $t$ is the product of the probabilities of the rules used to generate it.
- $P(s)$ – The probability of the string $s$ is the sum of the probabilities of the trees which have that string as their yield.

\[ P(s) = \sum_j P(s, t) \text{ where } t \text{ is a parse of } s = \sum_j P(t) \]

Tree and String Probabilities

- $s = \text{people fish tanks with rods}$
- $P(t_1) = 1.0 \times 0.7 \times 0.4 \times 0.5 \times 0.6 \times 0.7 \\
  \times 1.0 \times 0.2 \times 1.0 \times 0.7 \times 0.1$
  = 0.0008232

- $P(t_2) = 1.0 \times 0.7 \times 0.6 \times 0.5 \times 0.6 \times 0.2 \\
  \times 0.7 \times 1.0 \times 0.2 \times 1.0 \times 0.7 \times 0.1$
  = 0.00024696

- $P(s) = P(t_1) + P(t_2)$
  = 0.0008232 + 0.00024696
  = 0.00107016