PP Attachment

- See a man with a telescope
- Eat the pizza with a fork
- Eat the pizza with the anchovies

- PP attachments are a significant and frequent source of ambiguity
- Resolving PP attachments is difficult, in the worst case requiring world knowledge and general reasoning capabilities
- Tackling the PP attachment problem can give us insights into the more general parse selection problem
• **old boots and shoes**
• **a bank and warehouse guard**

• **Someone shot the servant of the actress who was on the balcony**
• **the lamps near the painting of the house that was damaged in the flood**

• Similar parsing sub-problems can be defined for coordination and relative clause attachments (which are just as hard in the worst case)
• Coordination, in particular, is still a general construction on which statistical parsers perform poorly

[ examples on the demo ]
• *Eat the pizza with a fork*

• *Eat the pizza with the anchovies*

• PCFGs only use *structural* probabilities (probabilities of CFG rules)

• Intuitively, the words are only generated at the leaves of the tree, “after” the crucial attachment decision has already been made

• The chosen analyses for the examples above will be the same (either verb attach or noun attach, depending on the rule probabilities)
A Real Example

- Pierre Vinken, 61 years old, joined the board as a nonexecutive director
- Did Vinken join as a nonexecutive director? (yes), or is it the board as a nonexecutive director? (no)
- NP-attach (incorrect): (joined ((the board) (as a nonexecutive director)))
- VP-attach (correct): ((joined (the board)) (as a nonexecutive director))
Focus on the 4 Headwords

• Pierre Vinken, 61 years old, joined the board as a nonexecutive director
• The decision can be made by considering the 4 headwords: join, board, as, director
• Examples like this can be extracted from the Penn Treebank (where 1 indicates noun attach, and 0 verb attach)
• 20,801 tuples for training; 4,039 for development; 3,097 for testing:

0 joined board as director
1 visited house on street
1 is chairman of N.V.
0 bring attention to problem
1 making paper for filters

...
A Probabilistic Formulation

\[ A_{\text{max}} = \arg \max_{A \in \{1,0\}} \hat{P}(A|V = v, N1 = n1, P = p, N2 = n2) \]

where \( A \) is the attachment site (1 for noun, 0 for verb), \( V \) is the verb, \( N1 \) is the object of the verb, \( P \) is the preposition, \( N2 \) is the object of preposition

- For example, for \textit{joined the board as a nonexecutive director}:

  \[ V = \textit{joined}, N1 = \textit{board}, P = \textit{as}, N2 = \textit{director} \]

- Simple algorithm:

  If \( \hat{P}(1|\text{joined, board, as, director}) > \hat{P}(0|\text{joined, board, as, director}) \)
  
  attach noun
  
  else

  attach verb
## Baselines and Upper Bounds

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always noun attach</td>
<td>59.0</td>
</tr>
<tr>
<td>Most likely based on preposition alone</td>
<td>72.2</td>
</tr>
<tr>
<td>Average human (4 head words only)</td>
<td>88.2</td>
</tr>
<tr>
<td>Average human (whole sentence)</td>
<td>93.2</td>
</tr>
</tbody>
</table>

from Ratnaparkhi et al. 1994
\[ \hat{P}(A|v, n_1, p, n_2) = \frac{f(A, v, n_1, p, n_2)}{f(v, n_1, p, n_2)} \]

- But what if \( f(v, n_1, p, n_2) = 0 \)? (Sound familiar?)
- Maybe we can use similar smoothing methods to what we have already seen for language modelling and tagging
If \( f(v, n1, p, n2) > 0 \) then \( \hat{P}(A|v, n1, p, n2) = \frac{f(A, v, n1, p, n2)}{f(v, n1, p, n2)} \)

else if \( f(triples) > 0 \) then \( \hat{P}(A|v, n1, p, n2) = \frac{f(A, triples)}{f(triples)} \)

else if \( f(pairs) > 0 \) then \( \hat{P}(A|v, n1, p, n2) = \frac{f(A, pairs)}{f(pairs)} \)

else if \( f(singleton) > 0 \) then \( \hat{P}(A|v, n1, p, n2) = \frac{f(A, singleton)}{f(singleton)} \)

• How should we break the 4-tuple into triples, pairs and singletons?
• The preposition is crucial, so always include that
A Backed-Off Model

If \( f(v, n_1, p, n_2) > 0 \) then
\[
\hat{P}(A|v, n_1, p, n_2) = \frac{f(A,v,n_1,p,n_2)}{f(v,n_1,p,n_2)}
\]

else if \( f(v, n_1, p) + f(v, p, n_2) + f(n_1, p, n_2) > 0 \) then
\[
\hat{P}(A|v, n_1, p, n_2) = \frac{f(A,v,n_1,p)+f(A,v,p,n_2)+f(A,n_1,p,n_2)}{f(v,n_1,p)+f(v,p,n_2)+f(n_1,p,n_2)}
\]

else if \( f(v, p) + f(n_1, p) + f(p, n_2) > 0 \) then
\[
\hat{P}(A|v, n_1, p, n_2) = \frac{f(A,v,p)+f(A,n_1,p)+f(A,p,n_2)}{f(v,p)+f(n_1,p)+f(p,n_2)}
\]

else if \( f(p) > 0 \) then
\[
\hat{P}(A|v, n_1, p, n_2) = \frac{f(A,p)}{f(p)}
\]

else \( \hat{P}(1|v, n_1, p, n_2) = 1.0 \) (i.e. default to noun attach)
### Results

<table>
<thead>
<tr>
<th>Stage</th>
<th>Total Number</th>
<th>Number Correct</th>
<th>Percent Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quad</td>
<td>148</td>
<td>134</td>
<td>90.5</td>
</tr>
<tr>
<td>Triples</td>
<td>764</td>
<td>688</td>
<td>90.1</td>
</tr>
<tr>
<td>Pairs</td>
<td>1965</td>
<td>1625</td>
<td>82.7</td>
</tr>
<tr>
<td>Singles</td>
<td>216</td>
<td>155</td>
<td>71.8</td>
</tr>
<tr>
<td>Default</td>
<td>4</td>
<td>4</td>
<td>100.0</td>
</tr>
<tr>
<td>Totals</td>
<td>3097</td>
<td>2606</td>
<td>84.1</td>
</tr>
</tbody>
</table>
• Perhaps surprisingly, using a higher value for the cutoff parameter reduced performance

• In other words, for this task, it is always better to use a non-zero count rather than back-off to a more general context (even if the count is 1)

• Morphological analysis (replacing years with YEAR, numbers with NUM, proper names with NAME, words with lemmas) gives only a small improvement (0.4%)
Other Approaches

- Some work on unsupervised PP attachment
  - based on the idea that some attachments are unambiguous; eg *the pizza with anchovies was tasty, eating with a fork is usual in the UK*

- Lots of work on using semantic information, e.g. from WordNet (if I know anchovies are a kind of food, fork is an implement . . .)
References


• Stetina and Nagoa (1997), Corpus based PP attachment ambiguity resolution with a semantic dictionary, Workshop on Very Large Corpora

• For more recent papers, do a search with PP attachment ambiguity