Parser Evaluation
View a parse as a set of labeled brackets / constituents

S(0,3)

NP(0,1)

PRP(0,1) (but standard evaluation does not count POS tags)

VP(1,3), VBD(1,2), NP(2,3), PRP(2,3)
**Parser Evaluation**

- **Precision**: number of correct predictions / number of predictions = $2/3$
- **Recall**: number of correct predictions / number of golds = $2/4$
- **F1**: harmonic mean of precision and recall = $(1/2 * ((2/4)^{-1} + (2/3)^{-1}))^{-1}$
  
  $= 0.57$ (closer to min)
Results

- Standard dataset for English: Penn Treebank (Marcus et al., 1993)
- “Vanilla” PCFG: ~71 F1
- Best PCFGs for English: ~90 F1
- State-of-the-art discriminative models (using unlabeled data): 95 F1
- Other languages: results vary widely depending on annotation + complexity of the grammar

Klein and Manning (2003)
Refining Generative Grammars
Language is not context-free: NPs in different contexts rewrite differently

[They]_{NP} received [the package of books]_{NP}
Vertical Markovization

Why is this a good idea?
Another way of doing lossless binarization:

- Equivalent to the original because of the chain rule of probability

\[
P(VBZ \ VBZ[VBZ] \mid VP): \text{write as } P(VBZ \mid VP) \text{ (slightly incorrect)}
\]

\[
P(VP \rightarrow VBZ \ NP \ PP \ PP) = P(VBZ \mid VP) \ P(NP \mid VP \ [VBZ]) \ \ P (PP \mid VP \ [VBZ \ NP]) \ P (PP \mid VP \ [VBZ \ NP \ PP])
\]
Changes amount of context remembered in binarization (h=∞: remember all)

In practice: always remember the *head* tag
First apply vertical Markovization, then binarize + apply horizontal
Annotating Trees

<table>
<thead>
<tr>
<th>Vertical Order</th>
<th>$h = 0$</th>
<th>$h = 1$</th>
<th>$h \leq 2$</th>
<th>$h = 2$</th>
<th>$h = \infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v = 1$ No annotation</td>
<td>71.27</td>
<td>72.5</td>
<td>73.46</td>
<td>72.96</td>
<td>72.62</td>
</tr>
<tr>
<td></td>
<td>(854)</td>
<td>(3119)</td>
<td>(3863)</td>
<td>(6207)</td>
<td>(9657)</td>
</tr>
<tr>
<td>$v \leq 2$ Sel. Parents</td>
<td>74.75</td>
<td>77.42</td>
<td>77.77</td>
<td>77.50</td>
<td>76.91</td>
</tr>
<tr>
<td></td>
<td>(2285)</td>
<td>(6564)</td>
<td>(7619)</td>
<td>(11398)</td>
<td>(14247)</td>
</tr>
<tr>
<td>$v = 2$ All Parents</td>
<td>74.68</td>
<td>77.42</td>
<td>77.81</td>
<td>77.50</td>
<td>76.81</td>
</tr>
<tr>
<td></td>
<td>(2984)</td>
<td>(7312)</td>
<td>(8367)</td>
<td>(12132)</td>
<td>(14666)</td>
</tr>
<tr>
<td>$v \leq 3$ Sel. GParents</td>
<td>76.50</td>
<td>78.59</td>
<td>79.07</td>
<td>78.97</td>
<td>78.54</td>
</tr>
<tr>
<td></td>
<td>(4943)</td>
<td>(12374)</td>
<td>(13627)</td>
<td>(19545)</td>
<td>(20123)</td>
</tr>
<tr>
<td>$v = 3$ All GParents</td>
<td>76.74</td>
<td>79.18</td>
<td>79.74</td>
<td>79.07</td>
<td>78.72</td>
</tr>
<tr>
<td></td>
<td>(7797)</td>
<td>(15740)</td>
<td>(16994)</td>
<td>(22886)</td>
<td>(22002)</td>
</tr>
</tbody>
</table>

Figure 2: Markovizations: $F_1$ and grammar size.

Klein and Manning (2003)
Tag Splits

- Can do some other specialized tag splits: e.g., sentential prepositions behave differently from other prepositions

- 79 F1 => 86.3 F1 using more tricks

Klein and Manning (2003)
Other Parsers
Lexicalized Parsers

- Even with parent annotation, these trees have the same rules. Need to use the words
Lexicalized Parsers

- Annotate each grammar symbol with its “head word”: most important word of that constituent

- Rules for identifying headwords (e.g., the last word of an NP before a preposition is typically the head)

- Collins and Charniak (late 90s): ~89 F1 with these