Not every NP expansion can fill every NP slot

- A grammar with symbols like “NP” won’t be context-free
- Statistically, treebank symbols represent overly strong conditional independence assumptions
Non-Independence I

- Independence assumptions are often too strong.

- Example: the expansion of an NP is highly dependent on the parent of the NP (i.e., subjects vs. objects).
- Also: the subject and object expansions are correlated!

Non-Independence II

- Who cares?
  - NB, HMMs, all make false assumptions!
  - For generation, consequences would be obvious.
  - For parsing, does it impact accuracy?

- Symptoms of overly strong assumptions:
  - Rewrites get used where they don’t belong.
  - Rewrites get used too often or too rarely.

In the PTB, this construction is for possessives
Breaking Up the Symbols

- We can relax independence assumptions by encoding dependencies into the PCFG symbols:

  ![Dependency Diagram]

  Parent annotation
  [Johnson 98]

  Marking possessive NPs

- What are the most useful “features” to encode?

Lexicalization

- Lexical heads important for certain classes of ambiguities (e.g., PP attachment):
  - Lexicalizing grammar creates a much larger grammar. (cf. next week)
    - Sophisticated smoothing needed
    - Smarter parsing algorithms
    - More data needed

- How necessary is lexicalization?
  - Bilexical vs. monolexical selection
  - Closed vs. open class lexicalization
Typical Experimental Setup

- Corpus: Penn Treebank, WSJ

  - Training: sections 02-21
  - Development: section 22 (here, first 20 files)
  - Test: section 23

- Accuracy – F1: harmonic mean of per-node labeled precision and recall.
- Here: also size – number of symbols in grammar.
  - Passive / complete symbols: NP, NP^S
  - Active / incomplete symbols: NP → NP CC

Horizontal Markovization

- Exhibit graphs showing the impact of horizontal Markov order on accuracy and symbol count for orders 1 and ∞.
Vertical Markovization

- Vertical Markov order: rewrites depend on past $k$ ancestor nodes. (cf. parent annotation)

Order 1

Order 2

Examples:
- Raw treebank: $v=1$, $h=\infty$
- Johnson 98: $v=2$, $h=\infty$
- Collins 99: $v=2$, $h=2$
- Best F1: $v=3$, $h=2v$

<table>
<thead>
<tr>
<th>Model</th>
<th>F1</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base: $v=h=2v$</td>
<td>77.8</td>
<td>7.5K</td>
</tr>
</tbody>
</table>
Unary Splits

- Problem: unary rewrites used to transmute categories so a high-probability rule can be used.
- Solution: Mark unary rewrite sites with -U

<table>
<thead>
<tr>
<th>Annotation</th>
<th>F1</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>77.8</td>
<td>7.5K</td>
</tr>
<tr>
<td>UNARY</td>
<td>78.3</td>
<td>8.0K</td>
</tr>
</tbody>
</table>

Tag Splits

- Problem: Treebank tags are too coarse.
- Example: Sentential, PP, and other prepositions are all marked IN.
- Partial Solution:
  - Subdivide the IN tag.

<table>
<thead>
<tr>
<th>Annotation</th>
<th>F1</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous</td>
<td>78.3</td>
<td>8.0K</td>
</tr>
<tr>
<td>SPLIT-IN</td>
<td>80.3</td>
<td>8.1K</td>
</tr>
</tbody>
</table>
**Other Tag Splits**

- **UNARY-DT**: mark demonstratives as DT^U (“the X” vs. “those”)
- **UNARY-RB**: mark phrasal adverbs as RB^U (“quickly” vs. “very”)
- **TAG-PA**: mark tags with non-canonical parents (“not” is an RB^VP)
- **SPLIT-AUX**: mark auxiliary verbs with –AUX [cf. Charniak 97]
- **SPLIT-CC**: separate “but” and “&” from other conjunctions
- **SPLIT-%**: “%” gets its own tag.

<table>
<thead>
<tr>
<th>F1</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>80.4</td>
<td>8.1K</td>
</tr>
<tr>
<td>80.5</td>
<td>8.1K</td>
</tr>
<tr>
<td>81.2</td>
<td>8.5K</td>
</tr>
<tr>
<td>81.6</td>
<td>9.0K</td>
</tr>
<tr>
<td>81.7</td>
<td>9.1K</td>
</tr>
<tr>
<td>81.8</td>
<td>9.3K</td>
</tr>
</tbody>
</table>

**A Fully Annotated (Unlex) Tree**

```
ROOT
  S'ROOT-v
    "S
      NP'S-B
        "S
          DT-U'NP
            This
          VBZ'BE'VP
            is
          NP'VP-B
            panic
            buying
      VP'S-VBF-v
        "S
```

Some Test Set Results

<table>
<thead>
<tr>
<th>Parser</th>
<th>LP</th>
<th>LR</th>
<th>F1</th>
<th>CB</th>
<th>0 CB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magerman 95</td>
<td>84.9</td>
<td>84.6</td>
<td>84.7</td>
<td>1.26</td>
<td>56.6</td>
</tr>
<tr>
<td>Collins 96</td>
<td>86.3</td>
<td>85.8</td>
<td>86.0</td>
<td>1.14</td>
<td>59.9</td>
</tr>
<tr>
<td>Unlexicalized</td>
<td>86.9</td>
<td>85.7</td>
<td>86.3</td>
<td>1.10</td>
<td>60.3</td>
</tr>
<tr>
<td>Charniak 97</td>
<td>87.4</td>
<td>87.5</td>
<td>87.4</td>
<td>1.00</td>
<td>62.1</td>
</tr>
<tr>
<td>Collins 99</td>
<td>88.7</td>
<td>88.6</td>
<td>88.6</td>
<td>0.90</td>
<td>67.1</td>
</tr>
</tbody>
</table>

- Beats “first generation” lexicalized parsers.
- Lots of room to improve – more complex models next.

The Game of Designing a Grammar

- Annotation refines base treebank symbols to improve statistical fit of the grammar
- Parent annotation [Johnson '98]
The Game of Designing a Grammar

- Annotation refines base treebank symbols to improve statistical fit of the grammar
  - Parent annotation [Johnson ’98]
  - Head lexicalization [Collins ’99, Charniak ’00]
Manual Annotation

- Manually split categories
  - NP: subject vs object
  - DT: determiners vs demonstratives
  - IN: sentential vs prepositional

- Advantages:
  - Fairly compact grammar
  - Linguistic motivations

- Disadvantages:
  - Performance leveled out
  - Manually annotated

<table>
<thead>
<tr>
<th>Model</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive Treebank Grammar</td>
<td>72.6</td>
</tr>
<tr>
<td>Klein &amp; Manning '03</td>
<td>86.3</td>
</tr>
</tbody>
</table>

Automatic Annotation Induction

- Advantages:
  - Automatically learned:
    - Label all nodes with latent variables.
    - Same number $k$ of subcategories for all categories.

- Disadvantages:
  - Grammar gets too large
  - Most categories are oversplit while others are undersplit.

<table>
<thead>
<tr>
<th>Model</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Klein &amp; Manning '03</td>
<td>86.3</td>
</tr>
<tr>
<td>Matsuzaki et al. '05</td>
<td>86.7</td>
</tr>
</tbody>
</table>
Learning Latent Annotations

EM algorithm:
- Brackets are known
- Base categories are known
- Only induce subcategories

Just like Forward-Backward for HMMs.

Refinement of the DT tag

DT
- the (0.50)
  - a (0.24)
    - The (0.08)
  - an (0.11)
- the (0.19)
- the (0.80)
  - The (0.15)
    - a (0.01)
- this (0.39)
  - that (0.28)
    - That (0.11)
  - all (0.19)
    - those (0.12)

DT-1 DT-2 DT-3 DT-4
Hierarchical refinement

Adaptive Splitting

- Want to split complex categories more
- Idea: split everything, roll back splits which were least useful
Adaptive Splitting

- Evaluate loss in likelihood from removing each split =

  \[ \text{Data likelihood with split reversed} \]

  \[ \text{Data likelihood with split} \]

- No loss in accuracy when 50% of the splits are reversed.

Adaptive Splitting Results

![Graph showing parsing accuracy (F1) vs. total number of grammar tasks.](image)

<table>
<thead>
<tr>
<th>Model</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous</td>
<td>88.4</td>
</tr>
<tr>
<td>With 50% Merging</td>
<td>89.5</td>
</tr>
</tbody>
</table>
## Final Results (Accuracy)

<table>
<thead>
<tr>
<th></th>
<th>≤ 40 words</th>
<th>all</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ENG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charniak &amp; Johnson ’05 (generative)</td>
<td>90.1</td>
<td>89.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Petrov and Klein 07</td>
<td>90.6</td>
<td>90.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GER</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dubey ’05</td>
<td>76.3</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Petrov and Klein 07</td>
<td>80.8</td>
<td>80.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chiang et al. ’02</td>
<td>80.0</td>
<td>76.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Petrov and Klein 07</td>
<td>86.3</td>
<td>83.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Final Results

<table>
<thead>
<tr>
<th>Parser</th>
<th>F1 ≤ 40 words</th>
<th>F1 all words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Klein &amp; Manning ’03</td>
<td>86.3</td>
<td>85.7</td>
</tr>
<tr>
<td>Matsuzaki et al. ’05</td>
<td>86.7</td>
<td>86.1</td>
</tr>
<tr>
<td>Collins ’99</td>
<td>88.6</td>
<td>88.2</td>
</tr>
<tr>
<td>Charniak &amp; Johnson ’05</td>
<td>90.1</td>
<td>89.6</td>
</tr>
<tr>
<td>Petrov et. al. 06</td>
<td>90.2</td>
<td>89.7</td>
</tr>
</tbody>
</table>
Learned Splits

- **Proper Nouns (NNP):**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NNP-12</td>
<td>John</td>
<td>Robert</td>
<td>James</td>
</tr>
<tr>
<td>NNP-2</td>
<td>J.</td>
<td>E.</td>
<td>L.</td>
</tr>
<tr>
<td>NNP-1</td>
<td>Bush</td>
<td>Noriega</td>
<td>Peters</td>
</tr>
<tr>
<td>NNP-15</td>
<td>New</td>
<td>San</td>
<td>Wall</td>
</tr>
<tr>
<td>NNP-3</td>
<td>York</td>
<td>Francisco</td>
<td>Street</td>
</tr>
</tbody>
</table>

- **Personal pronouns (PRP):**

<table>
<thead>
<tr>
<th>PRP-0</th>
<th>it</th>
<th>He</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRP-1</td>
<td>it</td>
<td>he</td>
<td>they</td>
</tr>
<tr>
<td>PRP-2</td>
<td>it</td>
<td>them</td>
<td>him</td>
</tr>
</tbody>
</table>

Learned Splits

- **Relative adverbs (RBR):**

<table>
<thead>
<tr>
<th>RBR-0</th>
<th>further</th>
<th>lower</th>
<th>higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBR-1</td>
<td>more</td>
<td>less</td>
<td>More</td>
</tr>
<tr>
<td>RBR-2</td>
<td>earlier</td>
<td>Earlier</td>
<td>later</td>
</tr>
</tbody>
</table>

- **Cardinal Numbers (CD):**

<table>
<thead>
<tr>
<th>CD-7</th>
<th>one</th>
<th>two</th>
<th>Three</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD-4</td>
<td>1989</td>
<td>1990</td>
<td>1988</td>
</tr>
<tr>
<td>CD-11</td>
<td>million</td>
<td>billion</td>
<td>trillion</td>
</tr>
<tr>
<td>CD-0</td>
<td>1</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>CD-3</td>
<td>1</td>
<td>30</td>
<td>31</td>
</tr>
<tr>
<td>CD-9</td>
<td>78</td>
<td>58</td>
<td>34</td>
</tr>
</tbody>
</table>
Coarse-to-Fine Parsing

[Goodman '97, Charniak&Johnson '05]

Coarse grammar

NP \ldots \ VP

Refined grammar

Parse

Prune

Treebank

Prune?

For each chart item $X[i,j]$, compute posterior probability:

$$\frac{P_{IN}(X, i, j) \cdot P_{OUT}(X, i, j)}{P_{IN}(root, 0, n)} < \text{threshold}$$

E.g. consider the span 5 to 12:

coarse:

refined:
Hierarchical Pruning

- Coarse:
- Split in two:
- Split in four:
- Split in eight:

Bracket Posteriors (after $G_0$)
Bracket Posteriors (Best Tree)

Parsing times

X-Bar=\(G_0\)  60 %
\(G_1\)  12 %
\(G_2\)  7 %
\(G_3\)  6 %
\(G_4\)  6 %
\(G_5\)  5 %
\(G=G_6\)  4 %
1621 min
111 min
35 min
15 min
(no search error)