PCFG Parsing, Evaluation, & Improvements

Ling 571
Deep Processing Techniques for NLP
January 24, 2011
Roadmap

- Parsing PCGFs:
  - Probabilistic CKY parsing

- Evaluation
  - Parseval

- Issues:
  - Positional and lexical independence assumptions

- Improvements:
  - Lexicalization: PLCFGs
Parsing Problem for PCFGs

- Select T such that:

\[ \hat{T}(S) = \arg \max_{T \text{ s.t. } S=\text{yield}(T)} P(T) \]

- String of words S is \( \text{yield} \) of parse tree over S
- Select tree that maximizes probability of parse

- Extend existing algorithms: CKY & Earley
  - Most modern PCFG parsers based on CKY
    - Augmented with probabilities
Probabilistic CKY

- Like regular CKY
  - Assume grammar in Chomsky Normal Form (CNF)
    - Productions:
      - $A \rightarrow B\ C$ or $A \rightarrow w$
  - Represent input with indices b/t words
    - E.g., Book$_0$ that$_1$ flight$_2$ through$_3$ Houston$_4$

- For input string length $n$ and non-terminals $V$
  - Cell[$i,j,A$] in $(n+1)\times(n+1)\times V$ matrix contains
    - Probability that constituent $A$ spans $[i,j]$
Probabilistic CKY Algorithm

function \text{PROBABILISTIC-CKY}(\text{words}, \text{grammar}) \text{ returns} \text{ most probable parse and its probability}

\text{for } j \leftarrow \text{from } 1 \text{ to } \text{LENGTH(words)} \text{ do}

\text{for all } \{ A \mid A \rightarrow \text{words}[j] \in \text{grammar} \}

\text{table}[j-1, j, A] \leftarrow P(A \rightarrow \text{words}[j])

\text{for } i \leftarrow \text{from } j - 2 \text{ downto } 0 \text{ do}

\text{for } k \leftarrow i + 1 \text{ to } j - 1 \text{ do}

\text{for all } \{ A \mid A \rightarrow BC \in \text{grammar},

\text{and } \text{table}[i, k, B] > 0 \text{ and } \text{table}[k, j, C] > 0 \}

\text{if } (\text{table}[i, j, A] < P(A \rightarrow BC) \times \text{table}[i, k, B] \times \text{table}[k, j, C]) \text{ then}

\text{table}[i, j, A] \leftarrow P(A \rightarrow BC) \times \text{table}[i, k, B] \times \text{table}[k, j, C]

\text{back}[i, j, A] \leftarrow \{k, B, C\}

\text{return } \text{BUILD-TREE}(\text{back}[1, \text{LENGTH(words)}, S]), \text{table}[1, \text{LENGTH(words)}, S]
### PCKY Grammar Segment

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$S$</td>
<td>$\rightarrow$</td>
<td>$NP$</td>
<td>$VP$</td>
<td>.80</td>
</tr>
<tr>
<td>$NP$</td>
<td>$\rightarrow$</td>
<td>$Det$</td>
<td>$N$</td>
<td>.30</td>
</tr>
<tr>
<td>$VP$</td>
<td>$\rightarrow$</td>
<td>$V$</td>
<td>$NP$</td>
<td>.20</td>
</tr>
<tr>
<td>$V$</td>
<td>$\rightarrow$</td>
<td>$includes$</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>$Det$</td>
<td>$\rightarrow$</td>
<td>$the$</td>
<td></td>
<td>.40</td>
</tr>
<tr>
<td>$Det$</td>
<td>$\rightarrow$</td>
<td>$a$</td>
<td></td>
<td>.40</td>
</tr>
<tr>
<td>$N$</td>
<td>$\rightarrow$</td>
<td>$meal$</td>
<td></td>
<td>.01</td>
</tr>
<tr>
<td>$N$</td>
<td>$\rightarrow$</td>
<td>$flight$</td>
<td></td>
<td>.02</td>
</tr>
</tbody>
</table>
PCKY Matrix:
The flight includes a meal

<table>
<thead>
<tr>
<th>Det: 0.4</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[0,1]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PCKY Matrix:
The flight includes a meal

<table>
<thead>
<tr>
<th>Det: 0.4</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[0,1]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N: 0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1,2]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**PCKY Matrix:**
*The flight includes a meal*

<table>
<thead>
<tr>
<th>Det: 0.4 [0,1]</th>
<th>NP: 0.3<em>0.4</em>0.2 = 0.0024 [0,2]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N: 0.2 [1,2]</td>
</tr>
</tbody>
</table>
PCKY Matrix:
The flight includes a meal

<table>
<thead>
<tr>
<th>Det: 0.4</th>
<th>NP: 0.3<em>0.4</em>0.2 = 0.0024</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[0,1]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N: 0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[1,2]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>V: 0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[2,3]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The flight includes a meal

### PCKY Matrix:

<table>
<thead>
<tr>
<th>Det: 0.4</th>
<th>NP: 0.3<em>0.4</em>0.2 = 0.0024</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[0,1]</td>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N: 0.2</td>
<td>[1,2]</td>
<td>[1,3]</td>
<td></td>
</tr>
<tr>
<td>V: 0.05</td>
<td>[2,3]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The flight includes a meal

<table>
<thead>
<tr>
<th>Det: 0.4</th>
<th>NP: 0.3<em>0.4</em>0.2 = 0.0024</th>
<th>[0,2]</th>
<th>[0,3]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0,1]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N: 0.2</td>
<td>[1,2]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[1,3]</td>
<td></td>
</tr>
<tr>
<td>[1,2]</td>
<td>V: 0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[2,3]</td>
<td></td>
</tr>
</tbody>
</table>
PCKY Matrix:
The flight includes a meal

<table>
<thead>
<tr>
<th>Det: 0.4</th>
<th>NP: (0.3 \times 0.4 \times 0.2 = 0.0024)</th>
<th>[0,2]</th>
<th>[0,3]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0,1]</td>
<td>N: 0.2</td>
<td>[1,2]</td>
<td>[1,3]</td>
</tr>
<tr>
<td></td>
<td>[1,2]</td>
<td>[2,3]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>V: 0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2,3]</td>
<td>Det: 0.4</td>
<td></td>
<td>[3,4]</td>
</tr>
</tbody>
</table>


PCKY Matrix:
The flight includes a meal

| Det: 0.4  | NP: 0.3*0.4*0.2 = .0024 |  |  
| [0,1]     | [0,2]                    | [0,3] |  
| N: 0.2    |  |  
| [1,2]     | V: 0.05                  |  | 
|          | [2,3]                    | [2,4] | Det: 0.4 
|          |                          |      | [3,4] |
PCKY Matrix:
The flight includes a meal

<table>
<thead>
<tr>
<th>Det: 0.4</th>
<th>NP: 0.3<em>0.4</em>0.2 = 0.0024</th>
<th>[0,3]</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[0,1]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N: 0.2</td>
<td></td>
<td>[1,3]</td>
<td>[1,4]</td>
<td></td>
</tr>
<tr>
<td>[1,2]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V: 0.05</td>
<td></td>
<td>[2,3]</td>
<td>[2,4]</td>
<td></td>
</tr>
<tr>
<td>[2,4]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Det: 0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[3,4]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PCKY Matrix:
The flight includes a meal

<table>
<thead>
<tr>
<th>Det: 0.4</th>
<th>NP: 0.3<em>0.4</em>0.2 = 0.0024</th>
<th>[0,3]</th>
<th>[0,4]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0,1]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N: 0.2</td>
<td>[1,3]</td>
<td>[1,4]</td>
</tr>
<tr>
<td>[1,2]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>V: 0.05</td>
<td>[2,3]</td>
<td>[2,4]</td>
</tr>
<tr>
<td>[2,4]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[3,4]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PCKY Matrix:
The flight includes a meal

<table>
<thead>
<tr>
<th>Det: 0.4</th>
<th>NP: 0.3*0.4*0.02 = 0.0024</th>
<th>[0,2]</th>
<th>[0,3]</th>
<th>[0,4]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0,1]</td>
<td></td>
<td>[1,2]</td>
<td>[1,3]</td>
<td>[1,4]</td>
</tr>
<tr>
<td>N: 0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1,2]</td>
<td></td>
<td>[2,3]</td>
<td>[2,4]</td>
<td></td>
</tr>
<tr>
<td>V: 0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2,3]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2,4]</td>
<td>Det: 0.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[3,4]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[3,4]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N: 0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[4,5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**PCKY Matrix:**
The flight includes a meal

<table>
<thead>
<tr>
<th>Det: 0.4 [0,1]</th>
<th>NP: 0.3<em>0.4</em>0.02 = 0.0024 [0,2]</th>
<th>[0,3]</th>
<th>[0,4]</th>
</tr>
</thead>
<tbody>
<tr>
<td>N: 0.2 [1,2]</td>
<td>[1,3]</td>
<td>[1,4]</td>
<td></td>
</tr>
<tr>
<td>V: 0.05</td>
<td>[2,3]</td>
<td>[2,4]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Det: 0.4 [3,4]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP: 0.3<em>0.4</em>0.01 = 0.0012 [3,5]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N: 0.01 [4,5]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PCKY Matrix: The flight includes a meal

<table>
<thead>
<tr>
<th>Det: 0.4</th>
<th>NP: 0.3<em>0.4</em>0.02 = 0.0024</th>
<th>[0,2]</th>
<th>[0,3]</th>
<th>[0,4]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0,1]</td>
<td>N: 0.2</td>
<td>[1,2]</td>
<td>[1,3]</td>
<td>[1,4]</td>
</tr>
<tr>
<td></td>
<td>V: 0.05</td>
<td>[2,3]</td>
<td>[2,4]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VP: 0.2<em>0.05</em>0.0012 = 0.000012</td>
</tr>
<tr>
<td></td>
<td>Det: 0.4</td>
<td>[3,4]</td>
<td></td>
<td>NP: 0.3<em>0.4</em>0.01 = 0.00012</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N: 0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[4,5]</td>
</tr>
</tbody>
</table>
PCKY Matrix:
The flight includes a meal

<table>
<thead>
<tr>
<th>Det: 0.4</th>
<th>NP: 0.3<em>0.4</em>0.02 = 0.0024</th>
<th>[0,1]</th>
<th>[0,2]</th>
<th>[0,3]</th>
<th>[0,4]</th>
<th>S: 0.000012* 0.0024 = 0.00000012</th>
<th>[0,5]</th>
</tr>
</thead>
<tbody>
<tr>
<td>N: 0.2</td>
<td>[1,2]</td>
<td>[1,3]</td>
<td>[1,4]</td>
<td>[1,5]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V: 0.05</td>
<td>[2,3]</td>
<td>[2,4]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Det: 0.4</td>
<td>[3,4]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N: 0.01</td>
<td>[4,5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Probabilistic Parser Development Paradigm

- Training:
  - (Large) Set of sentences with associated parses (Treebank)
    - E.g., Wall Street Journal section of Penn Treebank, sec 2-21
      - 39,830 sentences
    - Used to estimate rule probabilities
Probabilistic Parser Development Paradigm

- **Training:**
  - (Large) Set of sentences with associated parses (Treebank)
    - E.g., Wall Street Journal section of Penn Treebank, sec 2-21
    - 39,830 sentences
    - Used to estimate rule probabilities

- **Development (dev):**
  - (Small) Set of sentences with associated parses (WSJ, 22)
    - Used to tune/verify parser; check for overfitting, etc.
Probabilistic Parser Development Paradigm

- **Training:**
  - (Large) Set of sentences with associated parses (Treebank)
    - E.g., Wall Street Journal section of Penn Treebank, sec 2-21
    - 39,830 sentences
    - Used to estimate rule probabilities

- **Development (dev):**
  - (Small) Set of sentences with associated parses (WSJ, 22)
    - Used to tune/verify parser; check for overfitting, etc.

- **Test:**
  - (Small-med) Set of sentences w/parses (WSJ, 23)
    - 2416 sentences
    - Held out, used for final evaluation
Parser Evaluation

- Assume a ‘gold standard’ set of parses for test set
- How can we tell how good the parser is?
- How can we tell how good a parse is?
Parser Evaluation

- Assume a ‘gold standard’ set of parses for test set
- How can we tell how good the parser is?
- How can we tell how good a parse is?
  - Maximally strict: identical to ‘gold standard’
Parser Evaluation

- Assume a ‘gold standard’ set of parses for test set
- How can we tell how good the parser is?
- How can we tell how good a parse is?
  - Maximally strict: identical to ‘gold standard’
  - Partial credit:
Parser Evaluation

- Assume a ‘gold standard’ set of parses for test set
- How can we tell how good the parser is?
- How can we tell how good a parse is?
  - Maximally strict: identical to ‘gold standard’
  - Partial credit:
    - Constituents in output match those in reference
      - Same start point, end point, non-terminal symbol
Parseval

- How can we compute parse score from constituents?

- Multiple measures:
  - Labeled recall (LR):
    - \# of correct constituents in hyp. parse
    - \# of constituents in reference parse
Parseval

How can we compute parse score from constituents?

Multiple measures:

- Labeled recall (LR):
  - \# of correct constituents in \textit{hyp. parse}
  - \# of constituents in reference parse

- Labeled precision (LP):
  - \# of correct constituents in \textit{hyp. parse}
  - \# of total constituents in \textit{hyp. parse}
Parseval (cont’d)

• F-measure:
  • Combines precision and recall

\[
F_\beta = \frac{(\beta^2 + 1)PR}{\beta^2(P + R)}
\]

• F1-measure: \( \beta = 1 \)

\[
F_1 = \frac{2PR}{(P + R)}
\]

• Crossing-brackets:
  • # of constituents where reference parse has bracketing \(((A B) C)\) and hyp. has \((A (B C))\)
Precision and Recall

- Gold standard
  - (S (NP (A a) ) (VP (B b) (NP (C c)) (PP (D d))))

- Hypothesis
  - (S (NP (A a)) (VP (B b) (NP (C c) (PP (D d))))))
Precision and Recall

- Gold standard
  - \((S \ (NP \ (A \ a)) \ (VP \ (B \ b) \ (NP \ (C \ c)) \ (PP \ (D \ d))))\)

- Hypothesis
  - \((S \ (NP \ (A \ a)) \ (VP \ (B \ b) \ (NP \ (C \ c)) \ (PP \ (D \ d))))\)

- G: \(S(0,4) \ NP(0,1) \ VP(1,4) \ NP(2,3) \ PP(3,4)\)
Precision and Recall

- Gold standard
  - (S (NP (A a) ) (VP (B b) (NP (C c)) (PP (D d))))

- Hypothesis
  - (S (NP (A a)) (VP (B b) (NP (C c) (PP (D d)))))

- G: S(0,4) NP(0,1) VP (1,4) NP (2,3) PP(3,4)

- H: S(0,4) NP(0,1) VP (1,4) NP (2,4) PP(3,4)
Precision and Recall

- Gold standard
  - (S (NP (A a)) (VP (B b) (NP (C c)) (PP (D d))))

- Hypothesis
  - (S (NP (A a)) (VP (B b) (NP (C c) (PP (D d)))))

- G: S(0,4) NP(0,1) VP (1,4) NP (2,3) PP(3,4)
- H: S(0,4) NP(0,1) VP (1,4) NP (2,4) PP(3,4)
- LP: 4/5
Precision and Recall

- Gold standard
  - (S (NP (A a)) (VP (B b) (NP (C c)) (PP (D d))))
- Hypothesis
  - (S (NP (A a)) (VP (B b) (NP (C c)) (PP (D d))))
- G: S(0,4) NP(0,1) VP (1,4) NP (2,3) PP(3,4)
- H: S(0,4) NP(0,1) VP (1,4) NP (2,4) PP(3,4)
- LP: 4/5
- LR: 4/5
Precision and Recall

- Gold standard
  - (S (NP (A a) ) (VP (B b) (NP (C c)) (PP (D d)))))

- Hypothesis
  - (S (NP (A a)) (VP (B b) (NP (C c) (PP (D d)))))

- G: S(0,4) NP(0,1) VP (1,4) NP (2,3) PP(3,4)
- H: S(0,4) NP(0,1) VP (1,4) NP (2,4) PP(3,4)

- LP: 4/5
- LR: 4/5
- F1: 4/5
State-of-the-Art Parsing

- Parsers trained/tested on *Wall Street Journal* PTB
  - LR: 90%;
  - LP: 90%;
  - Crossing brackets: 1%

- Standard implementation of Parseval: `evalb`
Evaluation Issues

- Constituents?
Evaluation Issues

- Constituents?
- Other grammar formalisms
  - LFG, Dependency structure, ..
  - Require conversion to PTB format
Evaluation Issues

- Constituents?
  - Other grammar formalisms
    - LFG, Dependency structure, ..
    - Require conversion to PTB format

- Extrinsic evaluation
  - How well does this match semantics, etc?
Parser Issues

- PCFGs make many (unwarranted) independence assumptions
  - Structural Dependency
    - NP -> Pronoun: much more likely in subject position
  - Lexical Dependency
    - Verb subcategorization
    - Coordination ambiguity
Improving PCFGs: Structural Dependencies

- How can we capture Subject/Object asymmetry?
  - E.g., NP\textsubscript{subj} \rightarrow Pron vs NP\textsubscript{obj} \rightarrow Pron

- Parent annotation:
  - Annotate each node with parent in parse tree
    - E.g., NP\textsuperscript{S} vs NP\textsuperscript{VP}
    - Also annotate pre-terminals:
      - RB\textsuperscript{ADVP} vs RB\textsuperscript{VP}
      - IN\textsuperscript{SBAR} vs IN\textsuperscript{PP}

- Can also split rules on other conditions
Parent Annotation

(a) S
   NP  VP
   PRP VBD NP
       I need DT NN
          a flight

(b) S
   NP\$  VP\$
   PRP VBD NP\$VP
       I need DT NN
          a flight
Parent Annotation: Pre-terminals

```
VP'S
  TO
    to
    see
      IN
        if
          NN
            advertising
          NNS
            works
  VP'VP
    VB
    PP'VP
      IN
      NP'PP

VP'S
  TO'VP
    to
    see
      IN'SBAR
        if
          NP'S
            NN'NP
              advertising
            VBZ'VP
              works
  VP'VP
    VB'VP
    SBAR'VP
      S'SBAR
```
Parent Annotation

- Advantages:
  - Captures structural dependency in grammars
Parent Annotation

- Advantages:
  - Captures structural dependency in grammars

- Disadvantages:
  - Increases number of rules in grammar
Parent Annotation

- Advantages:
  - Captures structural dependency in grammars

- Disadvantages:
  - Increases number of rules in grammar
  - Decreases amount of training per rule
  - Strategies to search for optimal # of rules
Improving PCFGs: Lexical Dependencies

- Lexicalized rules:
  - Best known parsers: Collins, Charniak parsers
  - Each non-terminal annotated with its lexical head
    - E.g. verb with verb phrase, noun with noun phrase
  - Each rule must identify RHS element as head
    - Heads propagate up tree
  - Conceptually like adding 1 rule per head value

- VP(dumped) -> VBD(dumped)NP(sacks)PP(into)
- VP(dumped) -> VBD(dumped)NP(cats)PP(into)
Lexicalized PCFGs

• Also, add head tag to non-terminals
  • Head tag: Part-of-speech tag of head word
    • VP(dumped) -> VBD(dumped)NP(sacks)PP(into)
    • VP(dumped,VBD) -> VBD(dumped,VBD)NP(sacks,NNS) PP(into,IN)

• Two types of rules:
  • Lexical rules: pre-terminal -> word
    • Deterministic, probability 1
  • Internal rules: all other expansions
    • Must estimate probabilities
Internal Rules

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOP</td>
<td>$S(dumped,VBD)$</td>
</tr>
<tr>
<td>$S(dumped,VBD)$</td>
<td>$NP(workers,NNS)$ $VP(dumped,VBD)$</td>
</tr>
<tr>
<td>$NP(workers,NNS)$</td>
<td>$NNS(workers,NNS)$</td>
</tr>
<tr>
<td>$VP(dumped,VBD)$</td>
<td>$VBD(dumped,VBD)$ $NP(sacks,NNS)$ $PP(into,P)$</td>
</tr>
<tr>
<td>$PP(into,P)$</td>
<td>$P(into,P)$ $NP(bin,NN)$</td>
</tr>
<tr>
<td>$NP(bin,NN)$</td>
<td>$DT(a,DT)$ $NN(bin,NN)$</td>
</tr>
</tbody>
</table>

Lexical Rules

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$NNS(workers,NNS)$</td>
<td>$workers$</td>
</tr>
<tr>
<td>$VBD(dumped,VBD)$</td>
<td>$dumped$</td>
</tr>
<tr>
<td>$NNS(sacks,NNS)$</td>
<td>$sacks$</td>
</tr>
<tr>
<td>$P(into,P)$</td>
<td>$into$</td>
</tr>
<tr>
<td>$DT(a,DT)$</td>
<td>$a$</td>
</tr>
<tr>
<td>$NN(bin,NN)$</td>
<td>$bin$</td>
</tr>
</tbody>
</table>
PLCFGs

- Issue:
PLCFGs

- Issue: Too many rules
  - No way to find corpus with enough examples
PLCFGs

- Issue: Too many rules
  - No way to find corpus with enough examples

- (Partial) Solution: Independence assumed
  - Condition rule on
    - Category of LHS, head
  - Condition head on
    - Category of LHS and parent’s head

\[
P(T, S) = \prod_{n \in T} p(r(n) \mid n, h(n)) \times p(h(n) \mid n, h(m(n)))
\]
PLCFGs

- Issue: Too many rules
- No way to find corpus with enough examples
- (Partial) Solution: Independence assumed
  - Condition rule on
    - Category of LHS, head
  - Condition head on
    - Category of LHS and parent

\[
\prod \in T^n_n \cdot T^n_m \cdot T^n_h \cdot T^n_r \cdot S \cdot T \cdot P
\]
Disambiguation Example
Disambiguation Example

\[ P(\text{VP} \rightarrow \text{VBDNPPP} | \text{VP}, \text{dumped}) \]
\[ = \frac{C(\text{VP}(\text{dumped}) \rightarrow \text{VBDNPP})}{\sum_{\beta} C(\text{VP}(\text{dumped}) \rightarrow \beta)} \]
\[ = \frac{6}{9} = 0.67 \]

\[ p(\text{VP} \rightarrow \text{VBDNP} | \text{VP}, \text{dumped}) \]
\[ = \frac{C(\text{VP}(\text{dumped}) \rightarrow \text{VBDNP})}{\sum_{\beta} C(\text{VP}(\text{dumped}) \rightarrow \beta)} \]
\[ = \frac{0}{9} = 0 \]

\[ p(\text{in} | \text{PP}, \text{dumped}) \]
\[ = \frac{C(\text{X}(\text{dumped}) \rightarrow ... \text{PP}(\text{in})..)}{\sum_{\beta} C(\text{X}(\text{dumped}) \rightarrow ... \text{PP}...)} \]
\[ = \frac{2}{9} = 0.22 \]

\[ p(\text{in} | \text{PP}, \text{sacks}) \]
\[ = \frac{C(\text{X}(\text{sacks}) \rightarrow ... \text{PP}(\text{in})...)}{\sum_{\beta} C(\text{X}(\text{sacks}) \rightarrow ... \text{PP}...)} \]
\[ = \frac{0}{0} \]