Hierarchical Phrase-based Translation

Marzieh Razavi
Maryam Siahbani
Ravikiran Vadlapudi
Australia has diplomatic relations with North Korea, and is one of the few countries with formal diplomatic relations with North Korea. The text also indicates that Australia is one of the few countries with formal diplomatic relations with North Korea.
澳洲是与北韩有邦交的少数国家之一

Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi

Aozhou shi have dipl. rels. with N. Korea de shaoshu guojia zhiyi

< yu X you X, have X with X >
Aozhou is the few countries that have dep. rels. with North Korea.
Aozhou is one of the few countries that have dep. rels. with North Korea.
Synchronous CFG

\[ X \rightarrow < \gamma, \alpha, \sim > \]

- \( X \): non-terminal
- \( \gamma \): strings of terminals and non-terminals for source
- \( \alpha \): strings of terminals and non-terminals for target
- \( \sim \): 1-1 correspondence between non-terminals

\[ X \rightarrow < \text{yu X you X, have X with X} > \]
Rule Extraction
Rule Extraction

1. Identifying initial phrase pairs (similar to conventional phrase-based systems)

2. Extracting rules:
   a. Find phrases that contain other phrases
   b. Replace sub-phrases with non-terminals
Australia is one of the few countries that have dep. rels. with North Korea.

<table>
<thead>
<tr>
<th>Aozhou</th>
<th>is</th>
<th>one</th>
<th>of</th>
<th>the</th>
<th>few</th>
<th>countries</th>
<th>that</th>
<th>have</th>
<th>dep.</th>
<th>rels.</th>
<th>with</th>
<th>North Korea</th>
</tr>
</thead>
<tbody>
<tr>
<td>shi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beihan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>you</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bangjio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shaoshu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>guogia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zhiyi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Australia is one of the few countries that have dep. rels. with North Korea.

X → "yu X you X, have X with X" (6)

X → "Beiian, North Korea" (10)

X → "bangjio, dep. rels." (12)
Extracting Rules

<table>
<thead>
<tr>
<th></th>
<th>Australia</th>
<th>is</th>
<th>one</th>
<th>of</th>
<th>the</th>
<th>few</th>
<th>countries</th>
<th>that</th>
<th>have</th>
<th>dep.</th>
<th>rels.</th>
<th>with</th>
<th>North Korea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aozhou</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beihan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>you</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bangjio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shaoshu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>guogia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zhiyi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$X \overset{\rightarrow}{\sim} X \text{ de } X$, the $X$ that $X > (7)$
Australia is one of the few countries that have relations with North Korea.
Filtering the Grammar

• Limit the length of initial phrases to 10 words on either side.
• Limit the rules to five nonterminals plus terminals on the French side
• Rules can have at most two nonterminals
  – simplifies the decoder implementation.
• It is prohibited for nonterminals to be adjacent on the French side
  – major cause of spurious ambiguity
Other Rules

• Glue Rules: for dividing source side to chunks and translating one chunk at a time

\[ S \rightarrow < S_1 X_2, S_1 X_2 > \]  \hspace{1cm} (14)

\[ S \rightarrow < X_1 X_1 > \]  \hspace{1cm} (15)

• Entity Rules: for translating numbers, dates, ...

\[ X \rightarrow < X_1 \text{dunianlai, over the last } X_1 \text{years} > \]
\langle S_{10}, S_{11} \rangle

\Rightarrow \langle S_2 X_3, S_2 X_3 \rangle

\Rightarrow \langle S_4 X_5 X_3, S_4 X_5 X_3 \rangle

\Rightarrow \langle X_6 X_5 X_3, X_6 X_5 X_3 \rangle

\Rightarrow \langle Aozhou X_5 X_3, Australia X_5 X_3 \rangle

\Rightarrow \langle Aozhou shi X_3, Australia is X_3 \rangle

\Rightarrow \langle Aozhou shi X_7 zhiyi, Australia is one of X_7 \rangle

\Rightarrow \langle Aozhou shi X_5 de X_9 zhiyi, Australia is one of the X_9 that X_8 \rangle

\Rightarrow \langle Aozhou shi yu X_{11} you X_2 de X_9 zhiyi, Australia is one of the X_9 that have X_2 with X_{11} \rangle

\Rightarrow \langle Aozhou shi yu Beihan you X_2 de X_9 zhiyi, Australia is one of the X_9 that have X_2 with North Korea \rangle

\Rightarrow \langle Aozhou shi yu Beihan you bangjiao de X_9 zhiyi, Australia is one of the X_9 that have diplomatic relations with North Korea \rangle

\Rightarrow \langle Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi, Australia is one of the few countries that have diplomatic relations with North Korea \rangle
Model
Model

- General log-linear model over derivations D:

\[ P(D) \propto \prod_i \varphi_i(D)^{\lambda_i} \]

\[ \varphi_i(D) = \prod_{(X \to <\gamma,\alpha>) \in D} \varphi_i(X \to <\gamma,\alpha>) \]
Weighted Synchronous CFG

• Weights function over derivations D:

\[ w(D) = \prod_{(X \rightarrow \langle \gamma, \alpha \rangle) \in D} w(X \rightarrow \langle \gamma, \alpha \rangle) \]

• Weight for the rules:

\[ w(X \rightarrow \langle \gamma, \alpha \rangle) = \prod_{i \neq LM} \varphi_i(X \rightarrow \langle \gamma, \alpha \rangle)^{\lambda_i} \]

• Probability model:

\[ P(D) \propto P_{LM}(e)^{\lambda_{LM}} \times w(D) \]
Features

- \( P(\gamma | \alpha), P(\alpha | \gamma) \)
- Lexical weight \( P_w(\gamma | \alpha), P_w(\alpha | \gamma) \)
  - How well the words in \( \alpha \) translate the words in \( \gamma \)
- Language Model
- Extracted rules (with penalty \( \exp(-1) \))
- Glue rules (with penalty \( \exp(-1) \))
- Word penalty
- Dates, numbers, ....
Training
Training

- Estimate the parameters of phrase translation and lexical weighting:
  - Give a count 1 to each initial phrase pair occurrence
  - Distribute its weight uniformly among the rules obtained by subtracting sub-phrases from it
  - This distribution is considered as observed data
  - Use relative-frequency estimation to obtain $P(\gamma | \alpha), P(\alpha | \gamma)$

- Learn the parameters $\lambda_i$ of log-linear model:
  - MERT
Decoding
Basic Algorithm

• Objective

\[ \hat{e} = e \left( \arg \max_{D \text{ s.t. } f(D) = f} P(D) \right) \]

• Inference Rules

\[
\begin{align*}
Z \Rightarrow f_{i+1} & : w \\
[Z, i, i + 1] & : w \\
Z \Rightarrow XY & : w \\
[X, i, k] & : w_1 \\
[Y, k, j] & : w_2 \\
[Z, i, j] & : w_1w_2w
\end{align*}
\]
Basic Algorithm

Aozhou shi yu BeiHan you bangjiao de shaoshu guojia zhiyi

\[
X \rightarrow Aozhou \\
[X, 0, 1] : w_1
\]
Basic Algorithm

Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi

\[
\begin{align*}
X & \rightarrow Aozhou \\
[X, 0, 1] & : w_1 \\
X & \rightarrow shi \\
[X, 1, 2] & : w_3
\end{align*}
\]
Basic Algorithm

Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi

\[
\begin{align*}
X &\rightarrow Aozhou \\
[X,0,1] &: w_1 \\
X &\rightarrow shi \\
[X,1,2] &: w_3 \\
X &\rightarrow Beihon \\
[X,3,4] &: w_2 \\
X &\rightarrow bangjiao \\
[X,5,6] &: w_4 \\
X &\rightarrow shaoshu guojiao \\
[X,7,9] &: w_5
\end{align*}
\]
Basic Algorithm

Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi

\[
\begin{align*}
X & \rightarrow Aozhou \\
[X, 0, 1] : w_1 \\
X & \rightarrow shi \\
[X, 1, 2] : w_3 \\
X & \rightarrow Beihon \\
[X, 3, 4] : w_2 \\
X & \rightarrow bangjiao \\
[X, 5, 6] : w_4 \\
X & \rightarrow shaoshu guojia \\
[X, 7, 9] : w_5
\end{align*}
\]

\[
Z \rightarrow X zhiyi : w_6 \quad [X7, 9] : w_5 \\
\overline{[Z, 7, 11] : w_5 w_6}
\]
Basic Algorithm

Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi

\[
\begin{align*}
X &\rightarrow \text{Aozhou} \\
[X, 0, 1] & : w_1 \\
X &\rightarrow \text{shi} \\
[X, 1, 2] & : w_3 \\
X &\rightarrow \text{Beihon} \\
[X, 3, 4] & : w_2 \\
X &\rightarrow \text{bangjiao} \\
[X, 5, 6] & : w_4 \\
Z &\rightarrow X \text{zhiyi} : w_6 \\
[X, 7, 9] & : w_5 \\
[Z, 7, 11] & : w_5 w_6
\end{align*}
\]
Basic Algorithm

Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi

\[
\begin{align*}
X & \rightarrow Aozhou \\
[X, 0, 1] & : w_1 \\
X & \rightarrow shi \\
[X, 1, 2] & : w_3 \\
X & \rightarrow Beihon \\
[X, 3, 4] & : w_2 \\
X & \rightarrow bangjiao \\
[X, 5, 6] & : w_4 \\
Z & \rightarrow X zhiyi : w_6 \\
[X, 7, 9] & : w_5 \\
[Z, 7, 11] & : w_5w_6 \\
\end{align*}
\]

Z \rightarrow yu X_1 you X_2 : w_7 \quad [X_1, 3, 4] : w_2 \quad [X_1, 5, 6] : w_4 \\
[Z, 2, 6] : w_7w_2w_4
Basic Algorithm

Aozhou shì yù Beihàn yǒu bangjiao de shāoshū guojia zhīyi

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
</table>

\[
X \to Aozhou \\
\frac{\left[X, 0, 1\right]}{X \to shi} : w_1 \\
\frac{\left[X, 1, 2\right]}{X \to shi} : w_3
\]

\[
\begin{align*}
Z & \to yu X_1 you X_2 : w_7 & [X_1, 3, 4] : w_2 & [X_1, 5, 6] : w_4 \\
& \quad \quad \quad \quad [Z, 2, 6] : w_7w_2w_4
\end{align*}
\]

\[
\begin{align*}
Z & \to Xzhīyì : w_6 & [X, 7, 9] : w_5 \\
& \quad \quad \quad \quad [Z, 7, 11] : w_5w_6
\end{align*}
\]

Goal : \([S, 0, n]\)
**K-best Lists**

- Identify k-best derivations

- Used for Minimum error rate training

- Example: $L_1 = \{1,2,6,10\}$ and $L_2 = \{1,4,7\}$

\[
Z \rightarrow XY : w \quad \underbrace{[X, i, k] : w_1 \quad [Y, k, j] : w_2}_{L_1 \quad L_2}
\]

\[
[Z, i, j] : w_1w_2w
\]
$$Z \rightarrow XY : w \quad [X, i, k] : w_1 \quad [Y, k, j] : w_2$$

$$[Z, i, j] : w_1 w_2 w$$
\[ Z \rightarrow XY : w \]
\[ [X, i, k] : w_1 \]
\[ [Y, k, j] : w_2 \]
\[ [Z, i, j] : w_1 w_2 w \]
\[ Z \rightarrow XY : w \quad [X, i, k] : w_1 \quad [Y, k, j] : w_2 \]
\[ [Z, i, j] : w_1 w_2 w \]
\[ Z \to XY : w \quad [X, i, k] : w_1 \quad [Y, k, j] : w_2 \]
\[
\begin{array}{ccc}
& L2 & \\
L1 & 1 & 4 & 7 \\
 1 & 2 & 5 & 8 \\
 2 & 3 & 6 & \\
 6 & 7 & & \\
10 & & & \\
\end{array}
\]
Adding the Language Model

• Rescoring
  – Finding the k-best list using –LM parser
  – Rescoring the k-best list using LM
  – Linear in k
  – We may need to set k to be extremely high

• Intersection

• Cube Pruning
Intersection

\[ X \rightarrow \langle f^j_{i+1}, \alpha \rangle : \omega \]
\[ [X, i, j; q(\alpha)] : \omega p(\alpha) \]

\[ Z \rightarrow \langle f^i_{i+1} X f^j_{j_1+1}, \alpha \rangle : \omega \quad [X, i_1, j_1; e_1] : \omega_1 \]
\[ [Z, i, j; q(\alpha')] : \omega \omega_1 p(\alpha') \quad \alpha' = \alpha[e_1/X] \]

\[ Z \rightarrow \langle f^i_{i+1} X [\mathcal{D}] f^j_{j_1+1}, Y_2 f^j_{j_2+1}, \alpha \rangle : \omega \quad [X, i_1, j_1; e_1] : \omega_1 \quad [Y, i_2, j_2; e_2] : \omega_2 \]
\[ [Z, i, j; q(\alpha')] : \omega \omega_1 \omega_2 p(\alpha') \]
\[ \alpha' = \alpha[e_1/X [\mathcal{D}], e_2/Y_2] \]
Intersection

- Two function to correctly calculate the LM score of a sentence piecemeal

\[
p(a_1 \ldots a_l) = \prod_{m \leq i \leq l} P_{LM}(a_i | a_{i-m+1} \ldots a_{i-1}) \quad \text{if } l \geq m
\]

\[
q(a_1 \ldots a_l) = \begin{cases} 
  a_1 \ldots a_{m-1} \ast a_{l-m+1} \ldots a_l & \text{if } l \geq m \\
  a_1 \ldots a_l & \text{otherwise}
\end{cases}
\]

- \(p\) calculates LM probabilities for all the complete \(m\) grams
- \(q\) keeps the last and first \(m-1\) words of a string
Intersection

Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi

\[
X \rightarrow \langle Aozhou, Australia \rangle : w_1 \\
X, 0, 1 : w_1 p(Australia)
\]
## Intersection

Aozhou shi yu Bei Han you bangjiao de shaoshu guojia zhiyi

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
</table>

\[
\begin{align*}
X & \rightarrow \langle Aozhou, Australia \rangle : w_1 \\
[X, 0, 1] & : w_1 p(Australia) \\
X & \rightarrow \langle shi, is \rangle : w_3 \\
[X, 1, 2] & : w_3 p(is)
\end{align*}
\]
Intersection

Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi

\[
\begin{align*}
X &\rightarrow \langle Aozhou, Australia \rangle : w_1 \\
[X, 0, 1] &: w_1 p(Australia) \\
X &\rightarrow \langle shi, is \rangle : w_3 \\
[X, 1, 2] &: w_3 p(is) \\
X &\rightarrow \langle shaoshu guojio, few countries \rangle : w_5 \\
[X, 7, 9] &: w_5 p(few countries) \\
X &\rightarrow \langle Beihon, North Korea \rangle : w_2 \\
[X, 3, 4] &: w_2 p(North Korea) \\
X &\rightarrow \langle bangjiao, diplomatic relations \rangle : w_4 \\
[X, 5, 6] &: w_4 p(diplomatic relations)
\end{align*}
\]
Intersection

**Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi**

\[
\begin{align*}
X & \rightarrow \langle Aozhou, Australia \rangle : w_1 \\
& \quad [X, 0, 1] : w_1 p(Australia) \\
X & \rightarrow \langle shi, is \rangle : w_3 \\
& \quad [X, 1, 2] : w_3 p(is) \\
X & \rightarrow \langle shaoshu guojio, few countries \rangle : w_5 \\
& \quad [X, 7, 9] : w_5 p(few countries) \\
X & \rightarrow \langle Beihon, North Korea \rangle : w_2 \\
& \quad [X, 3, 4] : w_2 p(North Korea) \\
X & \rightarrow \langle bangjiao, diplomatic relations \rangle : w_4 \\
& \quad [X, 5, 6] : w_4 p(diplomatic relations) \\
Z & \rightarrow \langle yu X_1 you X_2, yu X_1 you X_2 \rangle : w_7 \\
& \quad [X_1, 3, 4, North Korea] : w'_2 \\
& \quad [X_2, 5, 6, diplomatic relations] : w'_4 \\
& \quad [Z, 2, 6, have dipl * with NK] : w_7 w'_2 w'_4 p(have dipl rels with NK)
\end{align*}
\]
Intersection

• Too slow in practice

• Pruning: for each span throw out items with score worse than:
  – the score of $b$th best item for that span
  – $\beta +$ the score of best item for that span
Cube Pruning

<table>
<thead>
<tr>
<th>X $\rightarrow$ $\langle$ cong $X_{\Box}$, from $X_{\Box}$ $\rangle$</th>
<th>1</th>
<th>2.1</th>
<th>5.1</th>
<th>8.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>X $\rightarrow$ $\langle$ cong $X_{\Box}$, from the $X_{\Box}$ $\rangle$</td>
<td>2</td>
<td>5.5</td>
<td>8.5</td>
<td>11.5</td>
</tr>
<tr>
<td>X $\rightarrow$ $\langle$ cong $X_{\Box}$, since $X_{\Box}$ $\rangle$</td>
<td>6</td>
<td>7.7</td>
<td>10.6</td>
<td>13.1</td>
</tr>
<tr>
<td>X $\rightarrow$ $\langle$ cong $X_{\Box}$, through $X_{\Box}$ $\rangle$</td>
<td>10</td>
<td>11.1</td>
<td>14.3</td>
<td>17.3</td>
</tr>
<tr>
<td>X, 6, 8; from the $\ast$ the scheme</td>
<td>5.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X, 5, 8; from the $\ast$ the plan</td>
<td>5.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X, 5, 8; since the $\ast$ the scheme</td>
<td>7.7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Cube Pruning

<table>
<thead>
<tr>
<th>X → 〈cong X, from X 〉</th>
<th>1</th>
<th>4</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.1</td>
<td>5.1</td>
<td></td>
</tr>
<tr>
<td>X → 〈cong X, from the X 〉</td>
<td>2</td>
<td>5.5</td>
<td>8.2</td>
</tr>
<tr>
<td>X → 〈cong X, since X 〉</td>
<td>6</td>
<td>8.5</td>
<td></td>
</tr>
<tr>
<td>X → 〈cong X, through X 〉</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table shows the scheme and project results for the given conditions.
Experiments
Experimental Results

- Comparing performances of decoding methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Settings</th>
<th>Time</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>rescore</td>
<td>$k = 10^4$</td>
<td>16</td>
<td>33.31</td>
</tr>
<tr>
<td>rescore</td>
<td>$k = 10^5$</td>
<td>139</td>
<td>33.33</td>
</tr>
<tr>
<td>intersect*</td>
<td></td>
<td>1455</td>
<td>37.09</td>
</tr>
<tr>
<td>cube prune</td>
<td>$\varepsilon = 0$</td>
<td>23</td>
<td>36.14</td>
</tr>
<tr>
<td>cube prune</td>
<td>$\varepsilon = 0.1$</td>
<td>35</td>
<td>36.77</td>
</tr>
<tr>
<td>cube prune</td>
<td>$\varepsilon = 0.2$</td>
<td>111</td>
<td>36.91</td>
</tr>
</tbody>
</table>
Experimental Results

• 2 baselines:
  – ATS
  – Hiero Monotone: same as Hiero except without any non-terminals on right hand side

<table>
<thead>
<tr>
<th>System</th>
<th>MT03</th>
<th>MT04</th>
<th>MT05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hiero Monotone</td>
<td>28.27 ± 1.03</td>
<td>28.83 ± 0.74</td>
<td>26.35 ± 0.92</td>
</tr>
<tr>
<td>ATS</td>
<td>30.84 ± 0.99</td>
<td>31.74 ± 0.73</td>
<td>30.50 ± 0.95</td>
</tr>
<tr>
<td>Hiero</td>
<td>33.72 ± 1.12</td>
<td>34.57 ± 0.82</td>
<td>31.79 ± 0.91</td>
</tr>
</tbody>
</table>
Questions ??