Modelling the Adjunct/Argument Distinction in Hierarchical Phrase-Based Translation

Sophie Arnoult and Khalil Sima’an

Institute for Logic, Language and Computation
University of Amsterdam

The Adjunct/Argument Distinction for Hiero

Minimally explain recursion in Hiero

- distinction is semantically driven
- adjunction is a central device for recursion
Interpretation of the Adjunct/Argument Distinction

A restrictive interpretation of the adjunct/argument distinction

- not modelling selectional preferences as in STAG
- adjuncts and arguments are interpreted as types in SCFG

Interpretation of adjuncts

- adjuncts as modifiers
- not only in semantic frames
Model

- Syntax-Augmented Machine Translation (SAMT)
  - labelled Hiero model
  - phrase labels derived from syntactic annotations through combinatory rules
- unlike SAMT
  - minimal labels
  - bilingual source/target annotations
- phrase-length constraint (10 tokens)
  - no labelled reordering at sentence level
Labelling procedure

Procedure

- adjunct/argument labels
- combinatory rules for phrase labels
- (bilingual) phrase-pair labels

Adjunct/Argument labels

- use dependency annotations
- map *modifier* and *punctuation* labels to adjuncts
Combinatory Rules for Phrase Labels

- derive phrase labels from adjunct (A) and argument (C) labels
- SAMT-like combinatory rules
- extension is minimal and reflects characteristics of adjunction

<table>
<thead>
<tr>
<th>phrase type</th>
<th>resulting label</th>
</tr>
</thead>
<tbody>
<tr>
<td>if constituent</td>
<td>A or C</td>
</tr>
<tr>
<td>else if constituent sequence</td>
<td></td>
</tr>
<tr>
<td>if all adjuncts</td>
<td>A</td>
</tr>
<tr>
<td>else</td>
<td>Cₜ</td>
</tr>
<tr>
<td>else if const. less sub constituents</td>
<td></td>
</tr>
<tr>
<td>if all adjuncts</td>
<td>A or C</td>
</tr>
<tr>
<td>else</td>
<td>Aᵢ or Cᵢ</td>
</tr>
<tr>
<td>else</td>
<td>P</td>
</tr>
</tbody>
</table>
Labelled Models

finally, arms fuel conflicts all over the world.

enfin, les armes alimentent les conflits de par le monde.
First Results

- French-English Europarl
- in-domain LM data, dev/test sets
- training with 200k sentence pairs

<table>
<thead>
<tr>
<th></th>
<th>labels</th>
<th>BLEU</th>
<th>METEOR</th>
<th>TER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>dev</td>
<td>test</td>
<td></td>
</tr>
<tr>
<td>Hiero</td>
<td>1</td>
<td>32.1</td>
<td>31.8</td>
<td>52.9</td>
</tr>
<tr>
<td>AA-Src</td>
<td>6</td>
<td>31.9</td>
<td>31.3</td>
<td>34.8</td>
</tr>
<tr>
<td>AA-Trg</td>
<td>6</td>
<td>32.0</td>
<td>31.6</td>
<td>34.9</td>
</tr>
<tr>
<td>AA-Bi</td>
<td>36</td>
<td>31.9</td>
<td>31.5</td>
<td>34.8</td>
</tr>
</tbody>
</table>
Relabelling by Clustering

- compare labels according to their \( lhs/rhs \) behaviour
- two-component distance
  - \( lhs \) distance
    \[
    d_{LHS} = \sum_{RHS} |\Delta_{LHS} P(rhs|lhs)|
    \]
  - \( rhs \) distance
    \[
    d_{cond_{RHS}} = \sum_{LHS} |\Delta_{RHS} P(lhs|rhs)|
    \]
    \[
    d_{joint_{RHS}} = \sum_{LHS} |\Delta_{RHS} P(lhs, rhs)|
    \]
- probabilities estimated from the dev-set AA-Bi grammar
- clustering stops at six clusters
Label Clusters

\[ d_{LHS} + d_{cond}^{cond} \]

\[ d_{LHS} + d_{joint}^{joint} \]

The Adjunct/Argument Distinction for Hiero
Sophie Arnout and Khalil Sima'an
## Results with Clustered Labels (1)

<table>
<thead>
<tr>
<th>labels</th>
<th>BLEU</th>
<th>METEOR</th>
<th>TER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dev</td>
<td>test</td>
<td>dev</td>
</tr>
<tr>
<td>AA-Bi</td>
<td>36</td>
<td>31.9</td>
<td>34.8</td>
</tr>
<tr>
<td></td>
<td>31.5</td>
<td>34.7</td>
<td>53.5</td>
</tr>
<tr>
<td>Cl-cond</td>
<td>6</td>
<td>31.8</td>
<td>34.8</td>
</tr>
<tr>
<td></td>
<td>▲▲</td>
<td>31.4</td>
<td>34.7</td>
</tr>
<tr>
<td>Cl-joint</td>
<td>6</td>
<td>31.9</td>
<td>34.9</td>
</tr>
<tr>
<td></td>
<td>31.8</td>
<td>34.8</td>
<td>▲▲ 53.3</td>
</tr>
</tbody>
</table>

The Adjunct/Argument Distinction for Hiero

Sophie Arnoult and Khalil Sima’an
### Results with Clustered Labels (2)

<table>
<thead>
<tr>
<th>labels</th>
<th>BLEU</th>
<th>METEOR</th>
<th>TER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dev</td>
<td>test</td>
<td>dev</td>
</tr>
<tr>
<td>Hiero</td>
<td>32.1</td>
<td>31.8</td>
<td>34.9</td>
</tr>
<tr>
<td>Cl-cond</td>
<td>31.8</td>
<td>31.4</td>
<td>34.8</td>
</tr>
<tr>
<td>Cl-joint</td>
<td>31.9</td>
<td><strong>31.8</strong></td>
<td>34.9</td>
</tr>
</tbody>
</table>
Future Work

- better method to reshape the bilingual-label set
  - clustering works, but only allows merging
- lift phrase-length constraint
  - reordering rules
  - swap for recursion constraint
- extend experimental set-up
  - other language pairs
Thank you.