Heuristic word alignment with parallel phrases

Maria Holmqvist
marho@ida.liu.se
Department of Computer and Information Science
Linköping University, Sweden

LREC May 19, 2010
Outline

- Word alignment with parallel phrases
- English-Swedish gold-standard word alignments
- Alignment experiments
- Comparison with Giza++
- Conclusions and future work
Word alignment

- Identify corresponding words in a text and its translation

I do **not think** it is **necessary** for **classic cars to be part of** the directive.

**Jag anser det inte nödvändigt** att **veteranbilar skall utgöra en del av** direktivet.

I – jag
I do not think – jag anser inte
I do not think it is necessary – jag anser det inte nödvändigt
...

- Applications:
  - Statistical machine translation
    - parallel segments (phrases) extracted from a word aligned parallel corpus
  - Bilingual dictionaries
  - Annotation projection
Word alignment with parallel phrases

- Parallel phrases from manually word aligned texts can be used to align new text.

- Extract all **parallel phrases** from a manually aligned parallel text
  - I do not think – jag anser inte – 0-0 2-2 3-1
  - do not think – anser inte – 1-1 2-0
  - ...

- Match parallel phrases to new sentence pairs

- If source and target phrase match the new sentence, add links

- Generalize phrases with parts-of-speech to increase recall

  - the european economy      den europeiska ekonomin      0-0 1-1 2-2
  - DET european economy      DET europeiska ekonomin      0-0 1-1 2-2
  - the A economy             den A ekonomin              0-0 1-1 2-2
  - DET A economy             DET A ekonomin              0-0 1-1 2-2
Word alignment with parallel phrases cont.

- in this N, I V  
i det N V jag  
0-0 1-1 2-2 4-4 5-3

| in this N, i V |
| In this sense, I am in agreement with Mr Sakellariou's proposals. |
| I det avseendet instämmer jag i Sakellarious förslag. |

- Heuristics for matching phrases to new sentences
  - Prefer longer phrases over shorter
    - more context -> better alignments
  - Prefer phrases with words over phrases with POS
    - DET european economy DET europeiska ekonomin
    - DET A N DET A N
  - Skip phrases that match more than once in the sentence
Manual word alignments

- English-Swedish Europarl corpus (Koehn, 2003)

<table>
<thead>
<tr>
<th>English</th>
<th>Swedish</th>
</tr>
</thead>
<tbody>
<tr>
<td>But we must support our tourism!</td>
<td>Men vi måste stödja vår turism!</td>
</tr>
<tr>
<td>These are events and situations that can not be tolerated.</td>
<td>Detta är händelser och situationer som vi inte kan acceptera.</td>
</tr>
<tr>
<td>Well in simple economic terms, Europe's culture adds to genuine prosperity.</td>
<td>Jo, i enkla ekonomiska termer främjar europeisk kultur verkligt välstånd.</td>
</tr>
</tbody>
</table>

- Training data set (1000 sentence pairs)
  - Extract parallel phrases
  - One annotator

- Reference set (200 sentence pairs)
  - Evaluation
  - Two annotators
  - Confidence labels for links
The English-Swedish reference word alignment

- Guidelines for English-Swedish reference alignment similar to Spanish-English guidelines (Lambert et al., 2005)
- Link types
  - Sure links
  - Possible links
  - Null links
- Two annotators
  - Agreement 85.8%
  - Alignments combined into the final reference alignment
  - Different link types -> Possible link
- The final reference alignment
  - 73% sure links
  - 27% possible links
Experiment 1: Generalized phrases

- Europarl English-Swedish training data
  - 1000 sentence pairs with manual alignments
  - parallel phrases 2-7 words were extracted from 900 sentence pairs
  - word alignment evaluated on 100 sentence pairs

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>phrases</td>
<td>92.25</td>
<td>16.85</td>
<td>28.50</td>
</tr>
<tr>
<td>generalized phrases</td>
<td>48.81</td>
<td>55.20</td>
<td>51.81</td>
</tr>
</tbody>
</table>

- Some generalized phrases propose incorrect word links
  - Ex. PREP N -> PREP N
    N -> N N
    of DET -> DET
Experiment 2: Constraining generalization

- Which generalized phrases produce correct word links?
- Thresholds for phrase length and generalization
  - P: Phrase length
  - L: Minimum phrase length for generalization
  - G: Max number of words generalized with POS

- P=1, L=3, G=1
  - Phrases of length 1 and up
  - Phrases have to be at least 3 words long to be generalized with POS
  - At most one word is generalized

<table>
<thead>
<tr>
<th>Thresholds (P-L-G)</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3-1</td>
<td>86.8</td>
<td>42.8</td>
<td>57.3</td>
</tr>
<tr>
<td>1-4-7</td>
<td>72.0</td>
<td>49.8</td>
<td>58.9</td>
</tr>
<tr>
<td>2-5-2</td>
<td>95.4</td>
<td>17.5</td>
<td>29.6</td>
</tr>
<tr>
<td>2-2-7</td>
<td>48.8</td>
<td>55.2</td>
<td>51.8</td>
</tr>
</tbody>
</table>
Experiment 3: Evaluating phrases on training data

- Apply phrases to the training data and collect
  - precision for each phrase
  - number of matches
- Perform word alignment with a subset of phrases
  - E.g., phrases with a training data precision above 95%

<table>
<thead>
<tr>
<th>Setting</th>
<th>Training precision threshold</th>
<th>Frequency threshold</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3-1</td>
<td>0.95</td>
<td>5</td>
<td>99.6</td>
<td>12.2</td>
</tr>
<tr>
<td>1-3-1</td>
<td>0.95</td>
<td>3</td>
<td>99.4</td>
<td>15.1</td>
</tr>
<tr>
<td>1-3-1</td>
<td>0.95</td>
<td>2</td>
<td>99.1</td>
<td>18.4</td>
</tr>
<tr>
<td>1-3-1</td>
<td>0.90</td>
<td>3</td>
<td>98.5</td>
<td>17.9</td>
</tr>
<tr>
<td>1-3-1</td>
<td>0.90</td>
<td>2</td>
<td>98.3</td>
<td>20.8</td>
</tr>
<tr>
<td>1-3-1</td>
<td>0.85</td>
<td>2</td>
<td>98.1</td>
<td>23.2</td>
</tr>
<tr>
<td>1-3-1</td>
<td>0.80</td>
<td>1</td>
<td>95.8</td>
<td>30.6</td>
</tr>
</tbody>
</table>
Creating final phrase-based alignments

- We have sets of reliable phrases
  - Generalization thresholds
  - Training data precision
  - Precision on the devtest set

- Alignment of test data
  - Combination of reliable phrases
  - Add linksets in sequence according to their precision on the devtest set

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F-mått</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combination1</td>
<td>95.85</td>
<td>28.27</td>
<td>43.66</td>
</tr>
<tr>
<td>Combination2</td>
<td>90.61</td>
<td>41.73</td>
<td>57.14</td>
</tr>
</tbody>
</table>
Giza++

- Giza++ (Och and Ney, 2003)
  - State-of-the-art system for statistical word alignment
  - Produces one-to-many alignments

- Method for symmetrization
  - **Intersection** – high precision
  - **Union** – high recall
  - **Grow-diag** – best AER

- Results for Giza++ trained on 700K sentences:

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
<th>AER</th>
</tr>
</thead>
<tbody>
<tr>
<td>intersect</td>
<td>94.77</td>
<td>57.05</td>
<td>71.22</td>
<td>16.31</td>
</tr>
<tr>
<td>union</td>
<td>70.09</td>
<td>77.17</td>
<td>73.46</td>
<td>21.77</td>
</tr>
<tr>
<td>grow-diag</td>
<td>82.35</td>
<td>73.30</td>
<td>77.56</td>
<td>15.46</td>
</tr>
</tbody>
</table>
# Results for Giza++ and phrase-based word alignment

Evaluation on test data (200 sentences)
- Precision is comparable
- Recall is lower than Giza++

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>AER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combination 1</td>
<td>95.8</td>
<td>28.3</td>
<td>45.7</td>
</tr>
<tr>
<td>Combination 2</td>
<td>90.6</td>
<td>41.7</td>
<td>31.5</td>
</tr>
<tr>
<td>Giza++ grow-diag 700K</td>
<td>82.3</td>
<td>73.3</td>
<td>15.5</td>
</tr>
<tr>
<td>Giza++ grow-diag 5K</td>
<td>71.4</td>
<td>62.0</td>
<td>26.6</td>
</tr>
<tr>
<td>Giza++ intersect 700K</td>
<td>94.8</td>
<td>57.1</td>
<td>16.3</td>
</tr>
<tr>
<td>Giza++ intersect 5K</td>
<td>93.3</td>
<td>42.8</td>
<td>28.7</td>
</tr>
</tbody>
</table>
A combination of Giza++ and phrase-based alignments

- Combine phrase based word alignment with statistical alignment
  - Add links from Giza++ for unaligned words
  - Results in improved AER

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>AER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giza++ 700k intersect</td>
<td>94.77</td>
<td>57.05</td>
<td>16.31</td>
</tr>
<tr>
<td>+ Combination1</td>
<td>93.41</td>
<td>60.10</td>
<td>14.81</td>
</tr>
<tr>
<td>Giza++ 700K grow-diag</td>
<td>82.35</td>
<td>73.30</td>
<td>15.46</td>
</tr>
<tr>
<td>+ Combination1</td>
<td>84.60</td>
<td>71.64</td>
<td>14.22</td>
</tr>
</tbody>
</table>
A combination of Giza++ and phrase-based alignments

- Larger improvements in AER for Giza++ trained on a small data set (5K)

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>AER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giza++ 5K intersect</td>
<td>93.3</td>
<td>42.8</td>
<td>28.7</td>
</tr>
<tr>
<td>+ Combination1</td>
<td>92.2</td>
<td>48.2</td>
<td>24.4</td>
</tr>
<tr>
<td>Giza++ 5K grow-diag</td>
<td>71.4</td>
<td>62.0</td>
<td>26.6</td>
</tr>
<tr>
<td>+ Combination1</td>
<td>75.1</td>
<td>62.2</td>
<td>23.9</td>
</tr>
</tbody>
</table>
Conclusions

- Word alignment with parallel phrases creates alignments with high precision (90-95%)
- Generalizing phrases with part-of-speech increased recall
- We have shown two methods to select reliable subsets of phrases that produce high-precision word alignments
- A combination of phrase-based word alignment and Giza++ produced a word alignment with lower AER than the best Giza++ alignment
Future work

- Other factors to find good phrase matches
  - relative position of matches in a sentence pair

- New language pairs

- Different merging strategy with Giza++. Perhaps during symmetrization.

- Better machine translation quality?
Thanks for listening!

Questions?