AER: Do we need to “improve” our alignments?

David Vilar, Maja Popović, Hermann Ney

International Workshop on Spoken Language Translation 2006
Kyoto, Japan
November 28, 2006

Human Language Technology and Pattern Recognition
Lehrstuhl für Informatik 6
Computer Science Department
RWTH Aachen University, Germany
Outline

1 Introduction
2 Related Work
3 Phrase-Based Translation
4 Tuple-Based Translation
5 Experimental Results
6 Discussion
1 Introduction

- Alignments are a key concept for statistical machine translation:
  - Represent the correspondence between the words of the source and target sentences.
- Most widely used alignment models: IBM1–IBM5 (single word-based models).
- State-of-the-art translation systems: phrase based models (or variations).
  - Most approaches make use of the Viterbi alignment in the training process.
  - Alignment used as an additional knowledge source.
It is expected that an increase in alignment quality leads to an increase in translation quality.

Alignment Error Rate (AER): Standard measure for alignment quality:

$$AER(S, P; A) = 1 - \frac{|A \cap S| + |A \cap P|}{|A| + |S|}.$$ 

We will present empirical evidence that a worse AER can result in an improvement in translation quality.
2 Related Work

[Fraser & Marcu 06] present an experimental study on the correlation of AER and BLEU.

They propose to use the F-measure directly as an indicator of alignment quality:

\[
F\text{-measure}(A, P, S, \alpha) = \frac{1}{\alpha \cdot \text{Precision}(A, P) + (1 - \alpha) \cdot \text{Recall}(A, S)}
\]

with

\[
\text{Precision}(A, P) = \frac{|A \cap P|}{|A|} \quad \text{Recall}(A, S) = \frac{|A \cap S|}{|S|}.
\]

We will also investigate this measure in this work.
3 Phrase-Based Translation

- Most important model(s) in a log-linear framework.
- Given a word-aligned parallel corpus, all phrases are extracted such that:
  1. all source words within the phrase are aligned only to target words within the phrase and
  2. all target words within the phrase are aligned only to source words within the phrase.
- More formally:

\[ \mathcal{BP}(f_1^J, e_1^I, A) = \{(f_{j+m}^j, e_{i+n}^i) \mid \forall (i', j') \in A : j \leq j' \leq j + m \iff i \leq i' \leq i + n\}. \]
Wie sieht es irgendwann morgens am Dienstag, dem sechsten aus?
how about sometime in the morning on Tuesday the sixth?
about sometime in the morning on Tuesday the sixth in?
4 Tuple-Based Translation

- Use joint probability of source and target string.
- Estimate an $m$-gram language model over the corpus of tuples $(\tilde{f}_k, \tilde{e}_k)$:

$$\hat{e}_1^I = \arg\max_{\tilde{e}_1^K, A, K} \prod_{k=1}^{K} p(\tilde{f}_k, \tilde{e}_k | \tilde{f}_{k-m}^{k-1}, \tilde{e}_{k-m}^{k-1}, A, K).$$

- RWTH system:
  - $\tilde{f}$ is always a single source word, $\tilde{e}$ is a phrase of 0 or more target words.
  - Use an alignment that is a function of target words.
- See also GIATI method proposed in [Casacuberta & Vidal 04].
¿Cuánto cuesta una habitación por semana?
the very beginning of May would suit me.
Alignment as a cost minimization problem:

- Introduce costs $c(i, j)$ of aligning source word $f_j$ to target word $e_i$.
- Define appropriate alignment constraints (full source or target sentence coverage, monotonicity, etc.).
- Find an alignment with minimal costs under these constraints.

In our case the alignment procedure is:

1. Find an alignment that is a function of the source words.
2. Reorder source sentence guided by this alignment.
3. Find an alignment that is a function of the target words.
The very beginning of May would suit me.

Mir wuerde sehr Anfang Mai passen.
the very beginning of May would suit me.
the very beginning of May would suit me.

sehr gut Anfang Mai wuerde passen mir.

the very beginning of May would suit me.

sehr | the very gut | $ Anfang | beginning
Mai | of May wuerde | would passen | suit mir | me .
the very beginning of May would suit me.

sehr gut Anfang Mai wuerde passen mir.

the very beginning of May would suit me.
5 Experimental Results

- Europarl Corpus (German to English)
- ACL2005 Shared Task

<table>
<thead>
<tr>
<th></th>
<th>German</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>751 088</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15 256 793</td>
<td>16 052 269</td>
</tr>
<tr>
<td></td>
<td>195 291</td>
<td>65 889</td>
</tr>
<tr>
<td>Test</td>
<td>2 000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>54 247</td>
<td>57 945</td>
</tr>
</tbody>
</table>

- Reference alignment on a subset of 508 randomly selected sentences.
- Both sure and possible alignments.
<table>
<thead>
<tr>
<th>System</th>
<th>Alignment</th>
<th>AER[%]</th>
<th>F[%]</th>
<th>BLEU[%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrase Based</td>
<td>Baseline</td>
<td>20.8</td>
<td>77.5</td>
<td>24.6</td>
</tr>
<tr>
<td></td>
<td>Phrases</td>
<td>24.2</td>
<td>71.8</td>
<td>24.8</td>
</tr>
<tr>
<td></td>
<td>Tuples</td>
<td>26.4</td>
<td>73.6</td>
<td>24.5</td>
</tr>
<tr>
<td>Tuple Based</td>
<td>Baseline</td>
<td>20.8</td>
<td>77.5</td>
<td>18.2</td>
</tr>
<tr>
<td></td>
<td>Phrases</td>
<td>24.2</td>
<td>71.8</td>
<td>14.8</td>
</tr>
<tr>
<td></td>
<td>Tuples</td>
<td>26.4</td>
<td>73.6</td>
<td>19.4</td>
</tr>
</tbody>
</table>
### Example for the Phrase-Based System

<table>
<thead>
<tr>
<th>Original</th>
<th>Es wird ein ganzes Kapitel über Wissenschaft, Gesellschaft und Bürger geben.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>It is a chapter on science, society and citizens.</td>
</tr>
<tr>
<td>Phrases</td>
<td>It will be a whole chapter on science, society and citizens.</td>
</tr>
<tr>
<td>Reference</td>
<td>There will be an entire chapter on science, society and the citizens.</td>
</tr>
</tbody>
</table>

### Example for the Tuple-Based System

<table>
<thead>
<tr>
<th>Original</th>
<th>Litauen verfügt über ein beträchtliches Potential für ein langfristiges Wirtschaftswachstum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>Has a considerable potential for a long-term Lithuania, although economic growth.</td>
</tr>
<tr>
<td>Tuples</td>
<td>Lithuania has a considerable potential for a long-term economic growth.</td>
</tr>
<tr>
<td>Reference</td>
<td>Lithuania has considerable potential for long-term economic growth.</td>
</tr>
</tbody>
</table>
6 Discussion

- AER and F-measure are appropriate for measuring alignment quality.
- “Inconsistency” between alignment models and translation models:
  - Improvements in alignment quality can lead to decreased translation performance (and vice-versa).
- [Ayan & Dorr 06] presents the “consistent phrase error rate”.
- Best solution would be that alignments remain a hidden variable in the statistical modelling process.
- Future work on alignments (oriented to machine translation) should always report results on translation quality.
References


