Source Phrase Segmentation and Translation for Japanese to English Translation Using Dependency Structure

Junki Matsuo, Kenichi Ohwada, and Mamoru Komachi
Graduate School of System Design, Tokyo Metropolitan University, Japan

We propose a linguistically motivated approach based on segmenting a source phrase using dependency structure and translating each phrase with PBSMT. This work presents the results of our method on Japanese-English translation and discusses potential improvements.

Algorithm
1. Our method creates a basic frame and dependent phrases. (Basic frame: broken line circle, Dependent phrases: solid line circle.)
2. The basic frame and the dependent phrases are translated by a decoder.
3. The anchor words of the basic frame are replaced with the translations of the corresponding dependent phrases. (Anchor words: the yellow under-lined words)

An example of the first method

An example of the second method

The difference between the first and second methods is not to replace a proposition that follows the predicate in the basic frame.

Because the translation of dependent phrases in our first method uses the language and translation models optimized for a sentence, our first method might not be able to use a model optimized for translating phrases.

Experimental Results

Error Analysis (in 100 sentences)

We investigate the reason why BLEU and RIBES fall off. The number of the pairs of input and reference that have different voices (active or passive) is 35.

We observed that translation of dependent phrases is the most frequent error type. It is because the language and the translation models are not optimized for translating phrases.

<table>
<thead>
<tr>
<th>Error types</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependency parsing</td>
<td>3</td>
</tr>
<tr>
<td>Translation of a basic frame</td>
<td>18</td>
</tr>
<tr>
<td>Translation of dependent phrases</td>
<td>46</td>
</tr>
<tr>
<td>Total (Each error may overlap)</td>
<td>57</td>
</tr>
</tbody>
</table>

Error Example

• Translation of the basic frame
  "雷撃比は等しい"
  ↓
  "The equal to lightning stroke ratio"
  → Ungrammatical Sentence

• Translation of the dependent phrases
  "解決を"
  ↓
  "DERSに"
  ↓
  "We solve"
  → "In DERS"
  ↓
  "Not an NP"
  → Missing context

Conclusion & Future Work

• Our finding is that our proposed methods have three problems: a dependency parsing, translation of a basic frame and translation of dependent phrases.
• We plan to optimize the language and the translation models for translating phrases.