Maximum Entropy Translation Model in Dependency-Based MT Framework

David Mareček, Martin Popel, Zdeněk Žabokrtský
Charles University in Prague, Institute of Formal and Applied Linguistics

Aim
• to improve translation quality by using context sensitive translation model (TM),
• exemplified on English-to-Czech MT.

Framework
• Analysis-Transfer-Synthesis
• Layers of language description adopted from the Prague Dependency Treebank:
  - Morphology
  - Surface dependency syntax
  - Deep dependency syntax
• Implemented in TectoMT

Resources
• 800 MW Czech monolingual corpus
• 60 MW parallel corpus CzEng 0.9 (parsed and aligned at the deep-syntactic level)

Translation Procedure

1. Morphological Analysis (Morce Tagger)
2. Surface-syntax Analysis (MST Parser)
3. Deep-syntax Analysis (Rules)
4. Transfer (TM+HMTM)
5. Surface-syntax Synthesis (Rules)
6. Morphological Synthesis (Rules + Stats)

Evaluation
• WMT 2010 test set (2489 sentences)
• Baseline and MaxEnt TMs also combined with Target Language Tree Model (TreeLM) using Hidden Markov Tree Model approach.
• Confirmed overlap between the contribution of MaxEnt TM and TreeLM.

<table>
<thead>
<tr>
<th>BLEU scores</th>
<th>No LM (just simple compatibility rules)</th>
<th>TreeLM (using Hidden Tree Markov Model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline TM</td>
<td>10.44</td>
<td>11.77</td>
</tr>
<tr>
<td>MaxEnt TM</td>
<td>11.77</td>
<td>12.58</td>
</tr>
<tr>
<td>improvement</td>
<td>+1.33</td>
<td>+0.81</td>
</tr>
</tbody>
</table>

Conclusion
MaxEnt-based translation models help to improve English-to-Czech translation quality in our dependency-based MT system.

However, he tried to find refuge in Brazil. Přesto se snažil najít útočiště v Brazílii.