Poor Man's Combination

- hybrid (rule-based/statistical) MT system
- transfer at a deep syntactic layer (t-layer)
- our combination: get an extra phrase table for Moses from TectoMT output

TectoMT Overview

- grammatical coherence, clause structure
- novel word forms (unseen in the parallel data)
- translations to better “glue” system outputs together
- not restricted to tokens in 1-best outputs, can use alternative table for Moses from TectoMT output
- transfer at a deep syntactic layer (t-layer)
- hybrid (rule-based/statistical) MT system
- reduction of modelling errors
- TectoMT provides many novel translations

Why TectoMT Helps

- TectoMT phrase table matches the test set → Moses can apply longer phrases
- better grammatical coherence
- search is simplified
- TectoMT provides many novel translations
- reduction of modelling errors

Poor Man's Combination

Parallel training data
Dev set (En)
Test set (En)
Baseline table
Synthetic table
Moses
Moses
CH0
CH1
Depfix
CH2

Constrained vs. Unconstrained

- monolingual: 44.3 vs. 392.3 million sentences
- parallel: 13.5 vs. 52.6 million sentence pairs

Language Models

long
- 7-gram LM on word forms
- mainly WMT monolingual data, individual years interpolated

big
- 4-gram LM on word forms
- use all available data

morph
- 10-gram LM on morphological tags

longmorph
- 15-gram LM on tags
- goal: capture sentential patterns

WMT Results

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU</th>
<th>TER</th>
<th>Manual</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH0</td>
<td>18.8</td>
<td>0.715</td>
<td>0.686</td>
</tr>
<tr>
<td>CH1</td>
<td>18.7</td>
<td>0.717</td>
<td></td>
</tr>
<tr>
<td>JHU-SMT</td>
<td>18.2</td>
<td>0.725</td>
<td>0.503</td>
</tr>
<tr>
<td>CH0</td>
<td>17.6</td>
<td>0.730</td>
<td></td>
</tr>
<tr>
<td>GOOGLE TRANSLATE</td>
<td>16.4</td>
<td>0.750</td>
<td>0.515</td>
</tr>
<tr>
<td>CU-TECTOMT</td>
<td>13.4</td>
<td>0.763</td>
<td>0.209</td>
</tr>
</tbody>
</table>

Chimera placed first among English-Czech MT systems in WMT for three years in a row.