The JHU Machine Translation Systems for WMT 2016
Shuoyang Ding, Kevin Duh, Huda Khayrallah, Philipp Koehn, and Matt Post

Phrase-Based System

- Baseline system from WMT 2015
- Och clusters for language model, OSM, reordering model, sparse features
- Huge language models on CommonCrawl data

<table>
<thead>
<tr>
<th>Language</th>
<th>Tokens</th>
<th>LM Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Czech</td>
<td>6.7 billion</td>
<td>13GB</td>
</tr>
<tr>
<td>German</td>
<td>65.2 billion</td>
<td>107GB</td>
</tr>
<tr>
<td>English</td>
<td>65.1 billion</td>
<td>89GB</td>
</tr>
<tr>
<td>Finnish</td>
<td>2.9 billion</td>
<td>8GB</td>
</tr>
<tr>
<td>Romanian</td>
<td>8.1 billion</td>
<td>13GB</td>
</tr>
<tr>
<td>Russian</td>
<td>23.3 billion</td>
<td>41GB</td>
</tr>
<tr>
<td>Turkish</td>
<td>11.9 billion</td>
<td>23GB</td>
</tr>
</tbody>
</table>
- Neural Network Joint Model

Domain Adapted Neural Language Model

- LM training data from different domains
  - Train model on all data, additional training with tuning set
  - Method 1, but only back-propagate through last layer of the network
  - Take the interpolation weights \( w_1, w_2, \ldots, w_n \) of the traditional language model interpolation. Monolingual data with word count \( c_1, c_2, \ldots, c_n \). Compute the normalized interpolation weights as follows:

  \[
  \tilde{w}_i = \frac{w_i}{c_i}
  \]

  Weight (by repeating) domain data with these weights

- Results on English–Romanian

<table>
<thead>
<tr>
<th>System</th>
<th>newsdev2016b</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>23.1</td>
</tr>
<tr>
<td>w/o untuned nplm on all data</td>
<td>23.5 (+ .4)</td>
</tr>
<tr>
<td>w/o untuned nplm on setimes2</td>
<td>23.2 (+ .1)</td>
</tr>
<tr>
<td>w/o all data nplm + method 1</td>
<td>23.4 (+ .3)</td>
</tr>
<tr>
<td>w/o all data nplm + method 2</td>
<td>23.8 (+ .7)</td>
</tr>
<tr>
<td>w/o all data nplm + method 3</td>
<td>24.0 (+ .9)</td>
</tr>
</tbody>
</table>

Neural Sequence Model Reranking

- Neural attention model (TensorFlow implementation)
- Reranked 50-best list
- Gains: Russian–English +.04, German–English +.10

Syntax-Based System

- Syntax-based system with Moses
  (German parser ParZu, English parser Berkeley)
- Hierarchical system: Joshua with Brown clusters

<table>
<thead>
<tr>
<th>Language Pair</th>
<th>Phrase</th>
<th>Syntax</th>
<th>Joshua</th>
</tr>
</thead>
<tbody>
<tr>
<td>English–Turkish</td>
<td>9.2</td>
<td>-</td>
<td>9.8</td>
</tr>
<tr>
<td>Turkish–English</td>
<td>12.9</td>
<td>13.9</td>
<td>-</td>
</tr>
<tr>
<td>English–Finnish</td>
<td>13.8</td>
<td>-</td>
<td>11.9</td>
</tr>
<tr>
<td>Finnish–English</td>
<td>19.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>English–Romanian</td>
<td>23.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Romanian–English</td>
<td>32.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>English–Russian</td>
<td>24.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Russian–English</td>
<td>27.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>English–Czech</td>
<td>23.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Czech–English</td>
<td>30.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>English–German</td>
<td>28.3</td>
<td>27.3</td>
<td>-</td>
</tr>
<tr>
<td>German–English</td>
<td>34.5</td>
<td>32.3</td>
<td>-</td>
</tr>
</tbody>
</table>

Morphological Segmentation

- Segment rare words
- Unsupervised segmentation: Morfessor & Byte Pair encoding
- Supervised segmentation: Chipmunk
- Turkish–English morphology results on newsdev2016b

<table>
<thead>
<tr>
<th>Method</th>
<th>Processing</th>
<th>Threshold</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>-</td>
<td>-</td>
<td>13.9</td>
</tr>
<tr>
<td>Byte-Pair preprocessing</td>
<td>-</td>
<td>-</td>
<td>13.7</td>
</tr>
<tr>
<td>Chipmunk</td>
<td>replace-rare</td>
<td>2</td>
<td>14.3</td>
</tr>
<tr>
<td>Chipmunk</td>
<td>replace-rare</td>
<td>10</td>
<td>14.9</td>
</tr>
<tr>
<td>Chipmunk</td>
<td>replace-rare</td>
<td>20</td>
<td>15.4</td>
</tr>
<tr>
<td>Chipmunk</td>
<td>replace-rare</td>
<td>20</td>
<td>14.7</td>
</tr>
<tr>
<td>Morfessor</td>
<td>replace-rare</td>
<td>0</td>
<td>13.5</td>
</tr>
<tr>
<td>Morfessor</td>
<td>replace-rare</td>
<td>2</td>
<td>13.7</td>
</tr>
<tr>
<td>Morfessor</td>
<td>replace-rare</td>
<td>5</td>
<td>14.0</td>
</tr>
<tr>
<td>Morfessor</td>
<td>replace-rare</td>
<td>10</td>
<td>14.1</td>
</tr>
<tr>
<td>Morfessor</td>
<td>replace-rare</td>
<td>20</td>
<td>14.2</td>
</tr>
</tbody>
</table>

Phrase-Based with Neural Joint Model

<table>
<thead>
<tr>
<th>Language Pair</th>
<th>Best 2015</th>
<th>Baseline</th>
<th>w/clusters</th>
<th>w/CC LM</th>
<th>w/both</th>
<th>w/NNJM</th>
<th>w/all&amp;tt100</th>
</tr>
</thead>
<tbody>
<tr>
<td>English–Turkish</td>
<td>-</td>
<td>7.8</td>
<td>8.2 +0.3</td>
<td>9.4 +1.6</td>
<td>8.9 +1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkish–English</td>
<td>-</td>
<td>14.0</td>
<td>14.3 +0.3</td>
<td>13.9 -0.1</td>
<td>14.1 +0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English–Finnish</td>
<td>15.5</td>
<td>11.9</td>
<td>12.6 +0.7</td>
<td>12.2 +0.3</td>
<td>12.9 +1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finnish–English</td>
<td>19.7</td>
<td>16.5</td>
<td>16.9 +0.4</td>
<td>16.4 -0.1</td>
<td>16.9 +0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English–Romanian</td>
<td>-</td>
<td>23.4</td>
<td>24.6 +1.2</td>
<td>23.4 +0.0</td>
<td>23.5 +0.1</td>
<td>23.7 +0.4</td>
<td>23.5 +0.1</td>
</tr>
<tr>
<td>Romanian–English</td>
<td>-</td>
<td>32.0</td>
<td>32.5 +0.5</td>
<td>32.5 +0.5</td>
<td>32.8 +0.8</td>
<td>32.0 +0.0</td>
<td>32.8 +0.8</td>
</tr>
<tr>
<td>English–Russian</td>
<td>24.3</td>
<td>23.9</td>
<td>25.0 +1.1</td>
<td>23.9 +0.0</td>
<td>24.9 +1.0</td>
<td>24.4 +0.5</td>
<td>25.2 +1.3</td>
</tr>
<tr>
<td>Russian–English</td>
<td>27.9</td>
<td>27.5</td>
<td>28.3 +0.7</td>
<td>28.1 +0.6</td>
<td>28.2 +0.7</td>
<td>27.8 +0.3</td>
<td>28.7 +1.2</td>
</tr>
<tr>
<td>English–Czech</td>
<td>18.8</td>
<td>18.2</td>
<td>19.2 +1.0</td>
<td>18.8 +0.6</td>
<td>19.6 +1.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czech–English</td>
<td>26.2</td>
<td>27.0</td>
<td>27.7 +0.6</td>
<td>27.7 +0.7</td>
<td>28.1 +1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English–German</td>
<td>24.9</td>
<td>22.7</td>
<td>23.0 +0.3</td>
<td>22.5 -0.2</td>
<td>22.7 +0.0</td>
<td>22.6 -0.1</td>
<td>22.9 +0.2</td>
</tr>
<tr>
<td>German–English</td>
<td>29.3</td>
<td>29.0</td>
<td>29.6 +0.6</td>
<td>29.6 +0.6</td>
<td>29.9 +0.9</td>
<td>29.6 +0.6</td>
<td>30.0 +1.0</td>
</tr>
</tbody>
</table>