Expanding the Language model in a low-resource hybrid MT system

G. Tambouratzis, S. Sofianopoulos, M. Vassiliou

Institute for Language & Speech Processing / Athena R.C.
Summary of Work

Motivation
How may an existing MT system be improved?

Possible Solution
Fuse different language models to improve translation accuracy
MT system

Hybrid MT system
The original MT system uses a syntactically-defined phrasal language model.

Planned upgrade
Supplement the phrasal model with n-gram language models.
Implementation

Sample a monolingual corpus to create a second TL-side language model.

Post-process MT output to identify problematic cases

Define hypotheses based on lemma- or token-type information

Consult n-gram statistics to determine necessary corrections to the MT output
## Results

<table>
<thead>
<tr>
<th>Number of sentences</th>
<th>Resources</th>
<th>Reference translations</th>
<th>Language pair</th>
<th>MT config.</th>
<th>BLEU</th>
<th>NIST</th>
<th>Meteor</th>
<th>TER</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>enrich.</td>
<td>1</td>
<td>EL–EN</td>
<td>Baseline</td>
<td>0.3008</td>
<td>6.541</td>
<td>0.3784</td>
<td>55.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$H_1$ to $H_4$</td>
<td>0.3059</td>
<td>6.569</td>
<td>0.3790</td>
<td>54.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$H_1$ to $H_5$</td>
<td><strong>0.3105</strong></td>
<td>6.593</td>
<td><strong>0.3791</strong></td>
<td>54.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$H_1$ to $H_6$</td>
<td>0.3096</td>
<td><strong>6.643</strong></td>
<td>0.3779</td>
<td><strong>54.64</strong></td>
</tr>
</tbody>
</table>