CMU System Combination in WMT 2011

Kenneth Heafield and Alon Lavie
{heafield,alavie}@cs.cmu.edu
kheafield.com/code

Search

Align
System 1
Projects that were revealed seemed promising.

System 2
The projects that have been proved seem promising.

Advance
System 1
Projects that were revealed seemed promising.

System 2
The projects that have been proved seem promising.

Exact matches
Snowball stems
WordNet synonyms
Unigram paraphrases

Pick from any system
Block aligned words
Skip stale words

System 1
Projects that were revealed seemed promising.

System 2
The projects that have been proved seem promising.

Partial Hypothesis
Projects that have been revealed

Score
Features

<table>
<thead>
<tr>
<th>Features</th>
<th>Exact Matches</th>
<th>Approximate Matches</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unigram</td>
<td>Bigram</td>
</tr>
<tr>
<td>System 1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>System 2</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Length 5
Language Model -23.8955, OOV 0

German Compounder
Recognize that "klar gemacht" aligns to "klargemacht".
Answer: segment → align → combine
Compounding:
1. Run every word through cdec’s segmenter
2. Count each form in normal text
3. Rejoin if the compound is more frequent
4. Score using sum of squares

Big LMs in 5 Languages
Data: Gigaword, News Crawl, News Commentary, UN, Europarl, Giga-FrEn, CzEng, and SMS (for ht-en)
Building: BigFatLM without pruning
Filtering: Union of sentence-filtered models
Inference: KenLM
github.com/jhclark and kheafield.com/code

Human Evaluation Results

Czech-English
Combination BLEU 29.85
Top Individual BLEU 29.99

German-English
Combination BLEU 26.21
Top Individual BLEU 24.30

Spanish-English
Combination BLEU 33.90
Top Individual BLEU 30.26

French-English
Combination BLEU 32.41
Top Individual BLEU 30.36

Clean Haitian-English
Combination BLEU 36.47
Top Individual BLEU 36.39

Raw Haitian-English
Combination BLEU 32.30
Top Individual BLEU 32.18

Discussion
Win on 8/10 tasks.
In Czech-English and German-English, online systems lead by > 1 BLEU. This work wins.
In French-English and Spanish-English, many systems scored near the top. RWTH and BBN perform better.
In Haitian-English, bm-i2r won by a large margin. Combination did little.
Out of English, there are fewer participants. Language-specific processing matters more.
Compared with other participants, this work does better when one system has a clear lead.

Supported by a NSF Graduate Research Fellowship and the DARPA GALE program.