QCRI at WMT12: Experiments in Spanish-English and German-English Machine Translation of News Text

Francisco Guzmán, Preslav Nakov, Ahmed Thabet, Stephan Vogel
Qatar Computing Research Institute, Qatar Foundation

1. Baseline System

- Train: News Commentary + Europarl;
- Dev: news2010;
- Test: news2011;
- Tokenize: split words containing a dash;
- Maximum sent. length: 100 tokens;
- Truecasing;
- Word alignments: IBM model 4 + grow-diag-final-and;
- Maximum phrase length: 7 tokens;
- Language model: 5-gram;
- Reordering: msd-bidirectional-fe;
- Detokenization: reconnecting words split on "-";
- Tuning: MERT.

2. Improvements

• Monotone at punctuation:	+0.15/+0.31
• MBR decoding:	+0.15/+0.18
• Kneser-Ney smoothing:	+0.18/+0.27
• Gigaword-by-year LM:	+0.37/+0.23
• Compound splitting:	------ /+0.22
• Tuning data selection:	+0.28/ ------
• Phrase table merging:	+0.42/ ------
• System combination:	+0.40/+0.20

3. Tuning Data Selection

$$s_f = \frac{1}{|G|} \sum_{g \in G} \texttt{sim}(f, g)$$

$$\texttt{sim}(f, g) = \texttt{2match}(f, g) * \texttt{lenpen}(f, g)$$

$$\texttt{lenpen}(f, g) = 3 - 4 * \texttt{sig}\left(\left[\frac{|f| - |g|}{\alpha}\right]^2\right)$$

devset	es-en
baseline (es:#6)	30.68
selection ($\alpha = 5$)	**30.94**
selection ($\alpha = 10$)	30.90

4. Phrase Table Combination

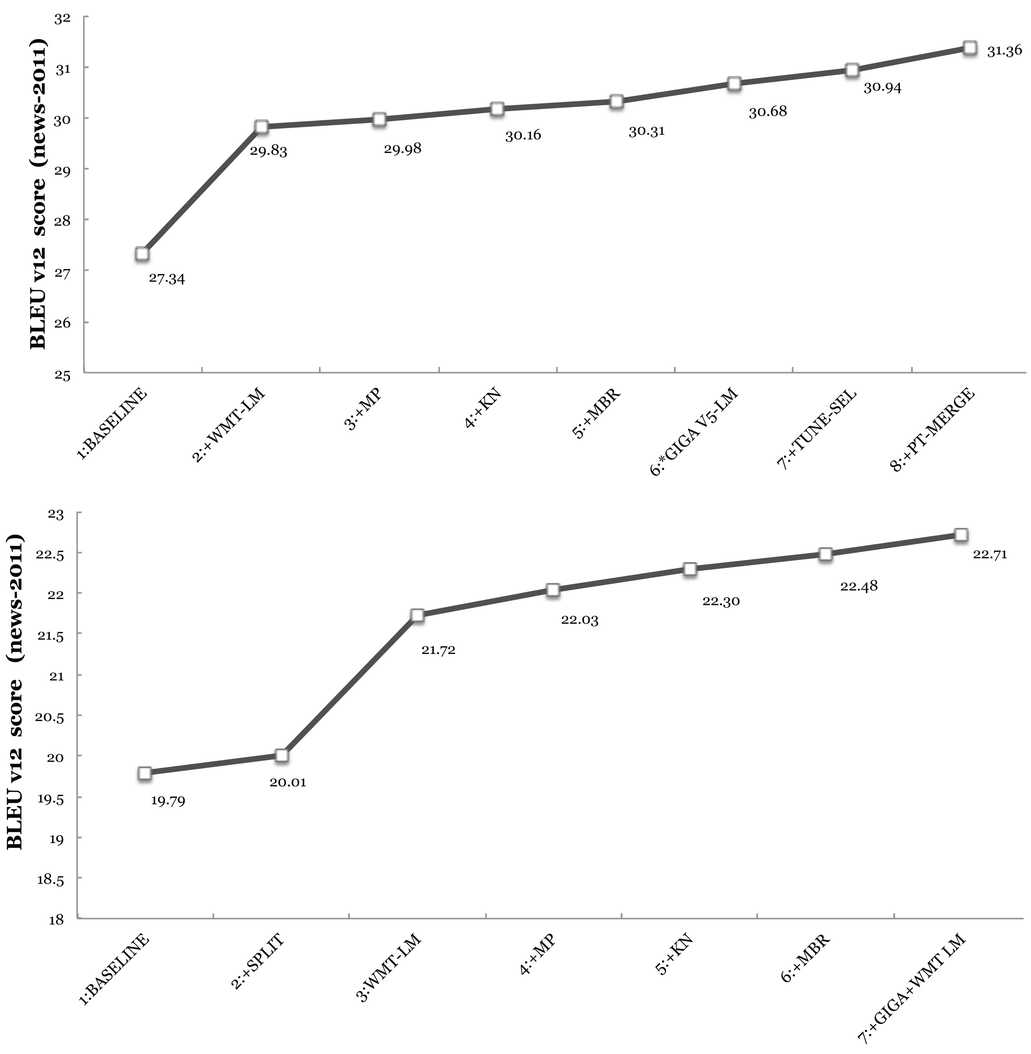
Two phrase tables:
(1) News Commentary
(2) Europarl

Merged using extra features:
F1 = 1 iff the pair came from (1);
F2 = 1 iff the pair came from (2);
F3 = 1 iff the pair came from (1) and (2);

Feature weights tuned using MERT.

	es-en
Baseline (es:#7)	30.94
Merging (1) News+EP with (2) UN	**31.36**

5. Experimental Results



7. Official Scores

	news2012 lower	news2012 cased
Spanish-English		
Primary	34.0	32.9
Secondary	33.6	32.5
German-English		
Primary	23.9	22.6
Secondary	23.6	22.4

Primary: best individual system
Secondary: system combination

8. Conclusion & Future Work

Ranking
- Second for Spanish-English
- In top-tier for German-English

Future work
- Data selection for tuning, training, and LM.