

Pattern Visualization for Machine Translation Output

Adam Lopez and Philip Resnik
University of Maryland, College Park

Problem: MT

Comparative Analysis

- Current methods such as BLEU are great at telling us, in aggregate, whether one system produces better output than another system
- They don't tell us anything about detailed competencies or weaknesses of one-system vs. another
- If we knew this, we might be able to more quickly improve our systems

Pattern-Based Analysis

- Looking at errors sentence by sentence will reveal lots of errors, but there will be a lot of variation and it will be difficult to make generalizations
- Our approach is to isolate recurrent patterns in the data that are common or interesting (e.g. one system is much better at them than another)

Method

- Tag all reference sentences
- Find correspondence between reference sentences and MT output sentences
- For all part-of-speech sequences in all reference sentence, determine the ratio of *recalled* sequences of words associated with that pattern (over total number of occurrences of that sequence)

Comparative Analysis

- We compared two systems to see if we could easily identify interesting patterns in their output
 - Hiero -- hierarchical phrase-based MT (Chiang, 2005)
 - Pharaoh -- standard phrase-based MT (Koehn, 2004)

Some Results

- First concrete evidence that Hiero is better at capturing long-distance reordering, noticeable in translation of nominal expressions
- Hiero baseline worse at number translation
- Hiero tends to delete possessive pronouns in certain cases
- Hiero often reformats the date in Arabic bylines