

Phrase-Based MT Summary

Phrase-based MT:

- is a log-linear model
- decomposes sentence into contiguous phrases (=MTU)
- assumes phrases relatively independent of each other
- search has two parts:
 - lookup of all relevant translation options
 - stack-based beam search, gradually expanding hypotheses

To train a PBMT system:

1. Align words.
2. Extract (and score) phrases consistent with word alignment.
3. Optimize weights (MERT).

Best implementation: Moses Decoder.