

Suggestions for Class Projects

Systematic study of phrase table generation heuristics

Use Moses pipeline, different heuristics, systematic changes of alignment density, alignment quality, corpus size to extract phrase pairs.

Comparison of Moses phrase-based and hierarchical decoding

Experiment with at least 2 language pairs; analyze the differences.

Compare GIZA and Berkeley aligner

Compare AER; impact on phrase table; impact on resulting translation.

Word alignment for language without word boundaries

Can standard work alignment models help to detect word/morpheme boundaries; experiment with simple (e.g. Spanish) and difficult language (e.g Inupiac)

Detect discontinuous phrase pairs

Analyze dis-contiguous phrases in hand aligned data; implement extension to PESA aligner to detect and align these dis-contiguous phrases.

Use Parts-of-Speech features to improve phrase alignment

Evaluate phrase alignment quality; implement extension to PESA aligner; evaluate impact on translation.

Sentence-level Confidence Estimation of MT quality

Possibly including identifying poor MT translations.

Tuning an MT system to different automatic metrics

Including the new METEOR-tuning, and comparing the outcomes.

Learning DNT (Do Not Translate) lists and incorporating them into Moses

Names, Acronyms, numbers, etc are often best carried over from the source to the target side without translating them.

Improving a Baseline Moses MT System

- Filtering out bad word-alignments
- Filtering the phrase-table
- Adding decoding features

Building a Hierarchical or Syntax-based MT system

For any language-pair.

Use Wordnet to extend phrase table