

# A Discriminative Lexicon for Translating to Morphologically Rich Languages

Alex Fraser and Kristina Toutanova

MTM

Sept 14th, 2013 - Charles University MTM

Leaders: Kristina Toutanova, Alex Fraser, Ondrej Bojar

- Decoding Team: Fabienne Braune and Ales Tamchyna
- Word Translation Extraction: Alexey Borisov and Sergey Gubanov
- Linguistic Feature Engineering: Marion Weller and Fran Tyers
- Experiment Pipeline and Stealth Project: Philipp Koehn

# Discriminative Word Lexicon

The goal is to modify a discriminative phrase lexicon (phrase-sense-disambiguation) to predict words.

- A phrase lexicon is a substitute for phrase-based  $p(e|f)$
- A discriminative **word lexicon** is a substitute for “lexical smoothing”  $p(e|f)$
- We are targeting English to Czech
- A Czech word is predicted using its alignment links to English (each Czech word is predicted more or less independently)

# Stealth Project

- We also implemented Minimal Translation Unit (MTU) extraction in the standard phrase-based framework
- The idea was to compare this with word prediction and phrase-sense-disambiguation
- We trained the standard phrase-sense-disambiguation pipeline on minimal units (so, a subset of the usual phrases)
- A short summary of the result: it is a possibly promising result showing issues that are conducive to further consideration
- 
- Baseline is interesting (French-to-English science set from the DAMT project, training size is 139K sentences)
- BLEU using phrases (score on tune): 29.70
- BLEU using MTUs (score on tune): 28.92

# Main Steps

- English to Czech baseline
- Defined lexicon lookup
  - ▶ A Czech word is predicted based on its aligned English words
  - ▶ Czech words can also be NULL-aligned
- Extract aligned English words for each Czech word in the parallel training data
  - We call this the "Aligned Group"
- Convert 15 position Czech tags to human readable form
- Convert feature extraction for classification from phrases to words

# Pipeline

- 1 Create annotated extraction file
- 2 Create translation table (with "Aligned Groups")
- 3 Run feature extraction
  - ▶ Baseline : source group, target word, absolute frequencies
  - ▶ Contrastive : morphologically rich annotation
- 4 Train classifier
- 5 Evaluate accuracy

# Used features in contrastive system

- source side features :
  - ▶ word, lemma, pos tag, number
  - ▶ features extracted from dependency trees :
    - ★ parent and grandparent nodes : lemma, pos tag
    - ★ leftmost child, rightmost child : lemma, pos, dependency relation
    - ★ left sister, right sister : lemma, pos
- target side features :
  - ▶ word, lemma
  - ▶ 15 features :
    - ★ part of speech
    - ★ number
    - ★ gender
    - ★ case, person, case, grade, negation,...

# Modifications to phrase-based setting I

- modified annotated extract files to handle "Aligned Groups"
  - ▶ sentence id
  - ▶ list of source spans
  - ▶ target span
  - ▶ annotated source group string
  - ▶ annotated target word string
- created a cept table to replace phrase-table
  - ▶ source cept
  - ▶ target cept
  - ▶ counts



## Modifications to phrase-based setting II

- modified feature extraction code to handle cepts instead of phrases
  - ▶ Extract from rich annotations from source group
  - ▶ Extract features from annotated target word
  - ▶ Extract word pairs from all alignment links
  - ▶ Extract features to count gaps in aligned group

## Reference sentence results on English-Czech

- Baseline
  - Source features - lexical identity of the aligned source group: *the\_actions*
  - Target features – identity of the target word together with its lemma and morphological tags: *akční\akční\AAIS4----1A----*
- Local context
  - Source features - source group, left word, right word
  - Target features - target

Feature set	Training Accuracy		Test Accuracy	
	Word+Factors	Word	Word+Factors	Word
Baseline	51.9	56.2	54.6	58.1
Local context	76.1	78.1	54.8	58.2

## Next steps

- Train VW for more iterations (we seem to be undertraining)
- Rename the discriminative word lexicon to something cooler (dwl is taken in Moses)
- Compare word-level, MTU-level and phrase-level prediction
  - ▶ Figure out how to compare these better (and run them on the same data)
  - ▶ Classification setup already works for word, MTU, phrase
  - ▶ Decoder for word-level is very close (MTU and phrase already implemented)
- ADD TARGET CONTEXT (to the left from the LM and phrase, to the right from the phrase!)
  - ▶ This requires moving classification into hypothesis extension (efficiency!!!)