# The RWTH Machine Translation System for IWSLT 2008

**David Vilar, Daniel Stein, Yuqi Zhang, Evgeny Matusov,
Arne Mause, Oliver Bender, Saab Mansour and Hermann Ney**

**IWSLT'08 21. October, 2008**

**Human Language Technology and Pattern Recognition
Lehrstuhl für Informatik 6
Computer Science Department
RWTH Aachen University, Germany**

# 1 Introduction

- Combination of phrase-based and hierarchical SMT systems
- Chinese-to-English and Arabic-to-English
- Investigated the effect of
  - different preprocessing techniques
  - reordering methods (including reordering of speech lattices)
  - syntax-based enhancements
- System ranked 6th in CE (all conditions) and 3rd in AE (all conditions)
- Combination of AE and CE outputs

# Outline

# 2 Translation Models

## 2.1 Phrase-based Model

- Well-known model
- Scores computed by relative frequencies
- Two different reordering models (depending on language pair)
  - IBM Reordering
  - Jump Reordering

# 2.2 Hierarchical Model

- Extension of the phrase-based model
- Allow for “gaps” in the phrases
- Formalized as a CF grammar (translation as parsing process)
- Example rules:

$$X \rightarrow \langle \text{中 } X^{\sim 0} \text{ 那个 } X^{\sim 1}, \text{ It's the } X^{\sim 1} \text{ in the } X^{\sim 0} \rangle$$
$$X \rightarrow \langle \text{也 要 } X^{\sim 0} \text{ 一些 } X^{\sim 1}, \text{ like to } X^{\sim 0} \text{ some } X^{\sim 1} \text{ too} \rangle$$

# 2.3 Common Models

- Word-based Lexicon Model
- Target Language Model (6-gram, Kneser-Ney discounting)
- Phrase Count Features
- Phrase Penalty
- Word Penalty

# 3 Extensions

## 3.1 Syntactical Features

- Extension for the hierarchical model
- Additional (soft) feature extracted at training time
- A rule is “syntactically consistent” if the “involved” original phrases correspond to yields of a syntax tree
- Done for source and target part independently
- Possibility of smoothing the “syntactic constraints”
- Detailed description in [Vilar et al. 2008]

# 3.2 Chunk-based Reordering for Chinese

- Reordering of the Chinese source sentence
- Syntactic chunk-level rules, automatically learned from the training data
- Reordered possibilities represented as $n$-best lists (with small $n$)
- Each reordering scored with the product of the probability of each of the rules
- Reordered training data added to the original data
- Detailed description in [Zhang et al. 2007]

# 3.3 Source Preprocessing

## Chinese

- Unigram segmenter obtains better results than ictclass
- LDC-like segmentation without text normalization
- Maximize the joint probability of all the words in the sentence
- Splitting long sentence pairs
- Detailed description in [Xu et al. 2008]

# Arabic

- Experiments with MADA and MorphTagger (POS-tagging tool)
- Three segmentation schemes
  - Splitting only the prefixes w+, l+, k+, b+, s+ (PRE)
  - Additionally splitting the determiner Al+ (PRE+DET)
  - Additionally splitting the pronominal suffixes (PREF+SUF)
- Tested normalizing Yaa and Alef
- Best results:
  - MorphTagger: PRE+SUF and no normalization
  - MADA: PRE and normalization

# 3.4 Translation of Speech Lattices

- Translation of word lattices including reordering
- Acoustic and source language model scores
- Cardinality synchronous search
  - Define cardinality in terms of “slots” (CN-like)
  - Allow for reordering without the over-generalization of CN
- Mapping from ASR vocabulary to MT vocabulary (segmentation)
- No improvements on this task (regretfully)
- Detailed description in [Matusov et al. 2008]

# 3.5 System Combination

- Approach used in last year's evaluation
- Build a confusion network for each sentence
  - Select one system as primary system
  - Align the single-best output of this system with the other hypotheses
  - Build a confusion network
  - Repeat with each system as primary
- The resulting confusion networks are joined into a word graph
- Weight with system specific factors and a trigram LM trained on the MT hypotheses
- Detailed description in [Matusov et al. 2006]

# 4 Experimental Results

- Arabic-to-English and Chinese-to-English translation directions
- Provided training data + HIT Corpus for CE
  - Selected sentences with 60% of the words in the IWSLT data
- Preprocessing of English:
  - Tokenization
  - Expansion of contractions
- GIZA++ for alignments
  - Tested different variants of word classes, model sequences and combination heuristics
- Optimized for BLEU on IWSLT 2004 eval data
- IWSLT 2005 eval data for system combination

# 4.1 Chinese-to-English

- BTEC Task

| CRR | | | | |
|---|---|---|---|---|
| System | BLEU | TER | WER | PER |
| System Combination | 46.1 | 37.7 | 43.9 | 39.4 |
| Phrase Based (PBT) | 42.5 | 36.6 | 45.3 | 40.6 |
| PBT + Chunk Reordering | 42.6 | 39.9 | 47.8 | 42.4 |
| PBT + New Segmentation | 44.3 | 40.3 | 47.3 | 42.0 |
| Hierarchical | 41.2 | 41.5 | 48.1 | 42.7 |
| Hierarchical + Syntax | 41.4 | 40.6 | 47.3 | 42.8 |
| **ASR** | | | | |
| System | BLEU | TER | WER | PER |
| System Combination | 39.7 | 42.5 | 49.6 | 44.5 |
| Phrase Based (PBT) | 37.3 | 41.2 | 50.0 | 45.1 |
| PBT + Chunk Reordering | 38.5 | 42.8 | 51.2 | 46.4 |
| Hierarchical | 31.6 | 49.6 | 56.5 | 49.5 |
| Hierarchical + Syntax | 36.6 | 44.1 | 51.4 | 47.0 |
| Lattices | 32.2 | 48.6 | 57.1 | 51.5 |

## Challenge Task

| CRR | | | | |
|---|---|---|---|---|
| System | BLEU | TER | WER | PER |
| System Combination | 39.1 | 40.7 | 48.3 | 44.1 |
| Phrase Based (PBT) | 32.1 | 42.7 | 51.9 | 47.8 |
| PBT + Chunk Reordering | 32.6 | 43.6 | 52.5 | 48.5 |
| PBT + New Segmentation | 37.2 | 41.8 | 49.3 | 44.5 |
| Hierarchical | 30.7 | 47.1 | 54.6 | 48.9 |
| Hierarchical + Syntax | 30.2 | 45.5 | 53.6 | 48.5 |
| **ASR** | | | | |
| System | BLEU | TER | WER | PER |
| System Combination | 34.3 | 43.6 | 51.1 | 46.1 |
| Phrase Based (PBT) | 27.8 | 46.0 | 55.4 | 51.1 |
| PBT + Chunk Reordering | 29.4 | 45.7 | 55.0 | 50.5 |
| Hierarchical | 26.4 | 51.0 | 59.2 | 51.9 |
| Hierarchical + Syntax | 30.2 | 45.6 | 53.7 | 48.6 |
| Lattices | 25.0 | 56.6 | 62.8 | 56.7 |

# 4.2 Arabic-to-English

| CRR | | | | |
|---|---|---|---|---|
| System | BLEU | TER | WER | PER |
| System Combination | 53.5 | 33.0 | 37.6 | 33.9 |
| PBT + MADA | 50.0 | 33.7 | 39.7 | 36.0 |
| PBT + MorphTagger | 51.8 | 33.8 | 38.1 | 33.9 |
| Hierarchical + MADA | 49.2 | 36.6 | 41.3 | 36.7 |
| Hierarchical + MorphTagger | 49.3 | 35.9 | 41.3 | 38.0 |
| **ASR** | | | | |
| System | BLEU | TER | WER | PER |
| System Combination | 44.5 | 37.6 | 43.4 | 39.9 |
| PBT + MADA | 42.6 | 38.2 | 45.3 | 41.7 |
| PBT + MorphTagger | 44.0 | 38.0 | 43.4 | 39.4 |
| Hierarchical + MADA | 41.3 | 42.1 | 47.7 | 42.7 |
| Hierarchical + MorphTagger | 41.3 | 40.7 | 47.2 | 43.9 |

- Note: bug in the hierarchical system (Corrected score for CRR: 54.1%)

# 4.3 Arabic&Chinese-to-English

- System combination of the best performing systems for both language pairs

| CRR | | | | |
|---|---|---|---|---|
| System | BLEU | TER | WER | PER |
| System Combination | 56.2 | 31.7 | 36.0 | 32.6 |

# 5 Conclusions

- Presented RWTH system for the IWSLT 2008 evaluation
- Combination of different statistical machine translation approaches
  - Phrase-based and hierarchical systems + extensions
- Combination of Arabic-to-English and Chinese-to-English systems increases performance