

## Overview

- ▶ RWTH participated in the Arabic-English BTEC task
- ▶ Different segmentation tools for Arabic
- ▶ Phrase-based and hierarchical SMT decoders
- ▶ System combination as final output

## Arabic Segmentation

- ▶ Off-the-shelf segmenters:
  - ▶ MADA - [Habash & Rambow 05] - SVM based classifier schemes: D1, D2, D3 and ATB
  - ▶ FST - [El Isbihani & Khadivi<sup>+</sup> 06] - rule based segmenter similar to ATB scheme plus additional splitting of “s” and “Al”
- ▶ Trained segmenters:
  - ▶ SVM (YAMCHA) - trained over the Arabic Treebank part 1
  - ▶ CRF (CRF++) - similar setup to the SVM
  - ▶ MorphTagger (SRILM) - HMM model selecting from Buckwalter Morphological Analyzer (BAMA)

## SMT Systems

- ▶ Training: IWSLT 2010 training data augmented with IWSLT03 and IWSLT07 test sets
- ▶ Development: IWSLT04 for tuning the translation system weights (MERT)
- ▶ Phrase based systems
  - ▶ PBT - [Zens & Ney 08] - in-house standard phrase-based system
  - ▶ Forced Alignments (FA) - [Wuebker & Mauser<sup>+</sup> 10] - retraining of the phrase table using a constrained decoder and leaving-one-out
- ▶ Hierarchical systems
  - ▶ JANE - [Vilar & Stein<sup>+</sup> 10] - open source hierarchical phrase-based system with standard setup
  - ▶ SYN - [Stein & Peitz<sup>+</sup> 10] - extend the set of non-terminals to include syntactic categories
  - ▶ POMS - [Vilar & Stein<sup>+</sup> 10] - as in SYN, but using automatic clustering methods to generate the categories

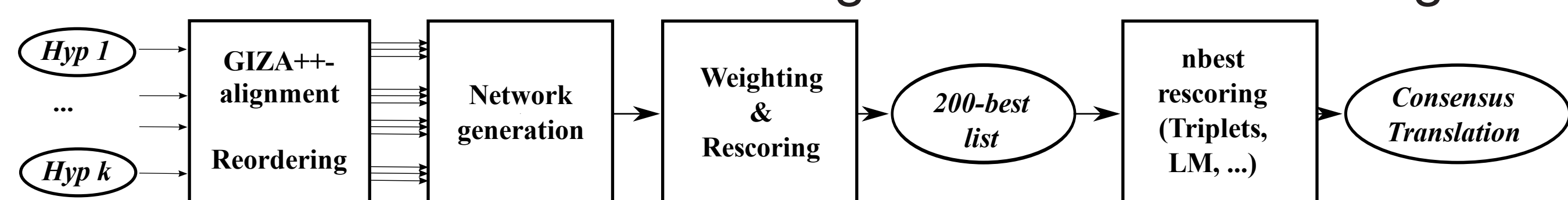
## Single System Results

System	PBT		FA		JANE		SYN		POMS	
	BLEU	TER	BLEU	TER	BLEU	TER	BLEU	TER	BLEU	TER
CRF	55.5	29.8	56.4	30.7	55.4	30.6	55.7	30.8	56.1	30.2
FST	54.5	30.7	55.9	30.3	55.3	30.2	54.4	31.2	56.0	29.4
MADA ATB	55.1	29.5	57.1	29.2	55.2	29.4	55.7	29.4	55.2	29.9
MADA D1	54.8	30.8	55.2	30.6	53.9	31.2	54.5	30.9	54.8	30.8
MADA D2	55.4	29.9	55.5	30.1	54.6	30.2	54.8	31.2	55.5	29.7
MADA D3	55.4	29.6	56.5	30.1	56.6	28.8	56.5	28.5	56.8	28.7
MorphTagger	56.5	29.2	55.8	30.1	57.1	29.4	56.6	29.2	57.5*	28.5*
SVM	56.1	29.7	55.9	30.0	56.6	28.9	55.4	30.3	54.9	29.5
TOK	55.5	30.1	54.8	30.3	53.0	32.4	52.7	32.5	53.4	32.3

- ▶ IWSLT 2008 lower case and with punctuation evaluation conditions (nocase+punc)
- ▶ high-degree of segmentation performs better than tokenization (but insignificant)
- ▶ hierarchical decoders perform better than phrase-based on average
- ▶ MorphTagger performs better for most of the decoders

## System Combination

- ▶ Based on [Leusch & Ney 10], state-of-the-art results in NIST 2009 and WMT2010
- ▶ Generates a lattice based on pairwise alignment of the different system outputs
- ▶ IWSLT05 test set to tune the weights for lattice rescoring



## Combination Experiments

- ▶ Schemes - different MADA schemes
- ▶ Methods - different segmentation methods
- ▶ SEG (combination) - different decoders using one segmenter
- ▶ Systems combi. - using best segmenter for each system
- ▶ Segmenters combi. - using best system for each segmenter
- ▶ Submission - best 15 systems then removal of “bad” ones
- ▶ results reported for IWSLT 2008 (nocase+punc)

## System Combination Results

System	BLEU	TER
PBT MADA Schemes	57.4	28.4
PBT Methods	58.0	28.3
FA MADA Schemes	57.8	28.8
FA Methods	58.1	28.7
JANE MADA Schemes	58.0	28.3
JANE Methods	59.0	27.8
SYN MADA Schemes	57.8	28.1
SYN Methods	57.7	28.2
POMS MADA Schemes	57.3	28.1
POMS Methods	59.2	27.4

⇒ Methods combination is better than Schemes combination

	SEG	BLEU	TER
Combi.	CRF	57.3	28.6
	FST	56.8	29.3
	MADA ATB	56.9	28.5
	MADA D1	55.2	30.2
	MADA D2	56.7	29.0
	MADA D3	58.1	27.5
	MorphTagger	58.1	28.2
	SVM	57.3	29.4
	TOK	54.8	31.1

⇒ combining different segmenters is better than combining different decoders

System	BLEU	TER
Systems combi.	58.2	27.8
Segmenters combi.	59.4	27.4
Submission	60.2	26.7