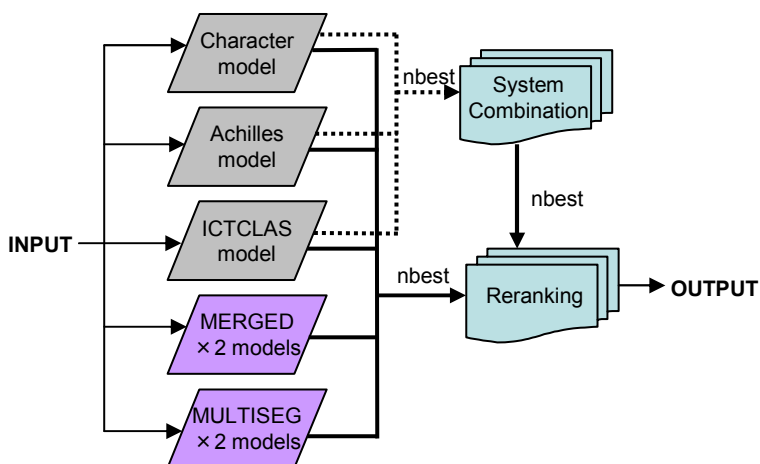


Chooi-Ling Goh, Taro Watanabe, Michael Paul, Andrew Finch, Eiichiro Sumita  
National Institute of Information and Communications Technology, Japan

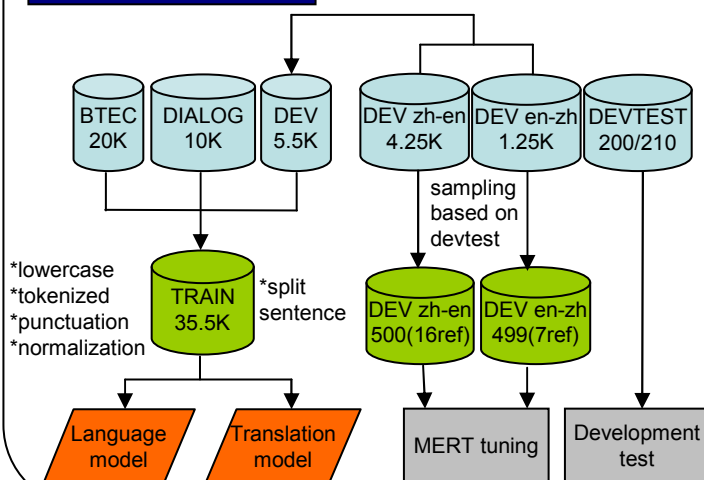
## Dialog task – Chinese-English

- ▶ Auto Chinese word segmentation bilingually
- ▶ Integration of multiple segmentation schemes (auto+standard)
- ▶ Decoder – CleopATRa – standard phrase-based SMT system
- ▶ Lattice-based system combination – Minimum Bayes-risk-Confusion Network framework
- ▶ SVM re-ranking – online large-margin learning for structured output based on MIRA – incorporated various global features such as words, POS tags, and contextual information

## Translation Flow



## Data Preparation



## Official Results

testset2010

System	English-Chinese		Chinese-English	
	ASR	CRR	ASR	CRR
Char	0.2743	0.3085	0.1850	0.1977
Merged	-	-	0.2013	0.2181
Multiseg	-	-	0.1835	0.2010
Syscomb 1-best	0.2874	<b>0.3172</b>	0.2099	<b>0.2332</b>
Syscomb 20-best	<b>0.2936</b>	-	<b>0.2103</b>	-
Rerank	0.2737	0.2999	0.1459	0.1616

## BTEC task – French-English

- ▶ CleopATRa – standard phrase-based SMT system
- ▶ Linparse – hierarchical phrase-based SMT system
- ▶ Word alignment with ITG-constrained
- ▶ Lattice-based system combination – Minimum Bayes-risk-Confusion Network framework
- ▶ Unknown word processing – replace unknown conjugated word form with known conjugated word form

## Data Preparation

Lowercase, tokenized, remove hyphens for French, split multiple sentences

## Official Results

System	IWSLT10	IWSLT09
CleopATRa	0.5146	0.5933
Linparse	0.5294	0.6108
Syscomb	<b>0.5395</b>	<b>0.6158</b>