

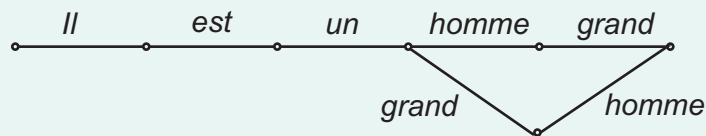
System Description

Baseline System

- State-of-the-art phrase-based SMT decoder
- Minimum Error Rate Training
- Phrase Table Smoothing
- 4-gram SRI language model
- Data TALK Task:
 - TM: EPPS + NC + UN + TED
 - LM: in addition Giga Corpus
- Data BTEC Task: provided data

Short-Range Reordering

- POS-based rules: NOM ADJ → ADJ NOM



Stemming

- French data stemmed to avoid sparse data issues
- English less inflectional
- LM sufficient
- generated by TreeTagger

Bilingual Language Model

- Introduce a bilingual language model using additional target word factors

Je	suis rentré	à la maison
I	went	home
I_Je	went_rentré_suis	home_la_maison

Adaptation

- exploit domain knowledge of in-domain data
- Translation model:
 - log-linear combination of general and in-domain relative frequencies
- Language model:
 - log-linear combination of language models

Results

- **POS-based reordering** models increase translation performance
- **Adaptation** needed to make best use of in-domain data
- **Bilingual Language Models** could improve the translation quality
- **Stemming** helps for tasks with small amounts of data

English-French TALK Task

System	Dev	Test
Baseline	27.39	22.70
+ Short-range Reordering	27.72	23.56
+ Language Model Adaptation	28.94	24.29
+ Translation Model Adaptation	29.57	24.68
+ UN corpus	29.40	24.78
+ 2-step Adaptation	29.62	24.67
+ Bilingual Language Model	30.08	24.95
+ Giga Language Model	30.39	36.34

French-English BTEC Task

	System	Test
1	Baseline	59.17
2	1 + source stems	60.83
3	1 + only alignment from stemmed source	60.73
4	2 + POS-based Reordering	60.55
5	2 + add. Postprocessing	61.51