NICT at WAT 2015

Chenchen Ding, Masao Utiyama, Sumita Eiichiro

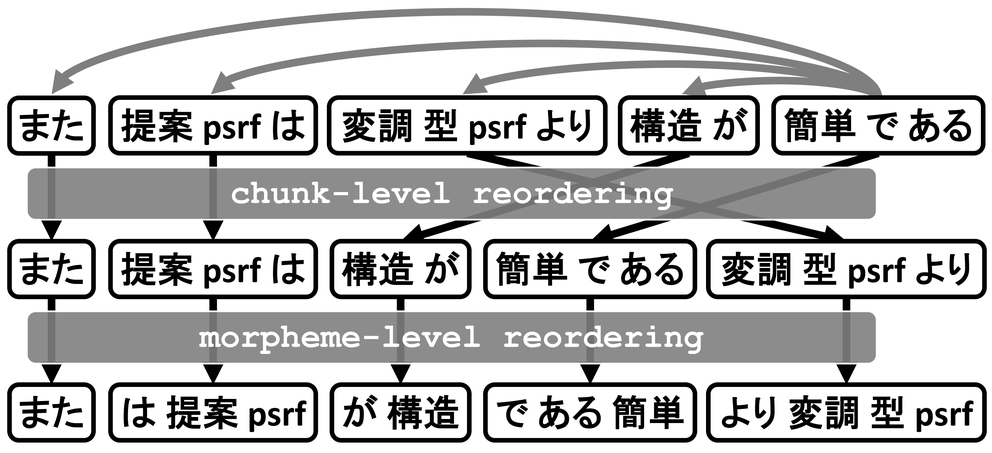
Multilingual Translation Laboratory
National Institute of Information and Communications Technology
{chenchen.ding, mutiyama, eiichiro.sumita}@nict.go.jp

Japanese-to-English (JE) SMT

Two pre-reordering approaches

- **REV-REO (morpheme-level)**
- **DEP-REO (chunk- / morpheme-level)**

REV-REO: J. Katz-Brown and M. Collins. 2008.
Syntactic reordering in preprocessing for Japanese → English translation: MIT system description for NTCIR-7 patent translation task.



DEP-REO: C. Ding, K. Sakanushi, H. Touji, and M. Yamamoto. 2015.
Inter-, Intra-, and Extra-chunk Pre-reordering for Statistical Japanese-to-English MT
(丁, 酒主, 通事, 山本. 2014. 統計的日英翻訳における依存構造に基づく事前並べ替えルール)

Korean-to-Japanese (KJ) SMT

Purely character-based

- **Unicode characters taken as basic units**
- **Nothing else**

Evaluation

SMT: Moses PB
LM: SRILM

JE translation

- **MeCab**
- **Cabocha**
- **mkn 5-gram**

KJ translation

- **max-ph-len=9**
- **mkn 9-gram**

JE trans.	BLEU	RIBES	HUM.
org. base.	18.45	.6451	-
our base.	18.09	.6397	-
REV-REO	18.96	.6845	+ 6.50
DEP-REO	18.98	.6599	+16.00

KJ trans.	BLEU	RIBES	HUM.
org. base.	69.73	.9408	-
our base.	70.92	.9427	+ 8.25
+post-proc.	71.11	.9429	+10.50

Conclusion

Even simple approaches work, as long as the properties of source and target languages considered

- **REV-REO for JE SMT may be taken as a new baseline because of its simplicity (and generality)**
- **Character-based KJ translation seems enough for SMT, while specific post-processing needed**