Fully Syntactic EBMT System of KYOTO Team in NTCIR-8

Toshiaki Nakazawa, Sadao Kurohashi
Graduate School of Informatics, Kyoto University

System Description
Input: whiskyはオオムギから製造される
Output: whisky is produced from barley

Alignment Model
IBM model
A photon is used for the photodetector

Dependency Relation Probability
Parent-child
E_c
Grandparent-child
F_p
Inverted parent-child
F'_c

Model Training
Step 1: Estimate word translation prob.
- Initialize dependency relation prob.

Phrase Translation Probability
A^c_f = 2, A^p_f = 3, A^c_p = 0
A^p_c = 0, A^c_p = 1, A^p_c = 2
p(f | a, e) = \prod p(f_i | E_{p_i})
= p(f_1 | E_1) \cdot p(f_2 | E_2) \cdot p(f_3 | NULL)
p(e | a, f) = p(E_1 | NULL) \cdot p(E_2 | F_1) \cdot p(E_2 | F_2)

Possible Phrase Generation
Initial alignment
Swap
Add
Extend
Reject

Translation Evaluation Results
Direction Training Corpus Parsing Accuracy Kyoto-U RBMT Moses
BLEU Adeq. BLEU Adeq. BLEU Adeq.
Ja->En About 1M 90% (En) 90% (Ja) 19.2 3.60 11.6 3.64 20.9 3.35
En->Ja 21.2 --- 10.3 --- 21.4 ---
Ja->Zh About 6OOK 70% (Zh) 90% (Ja) 14.6 --- 8.9 --- 17.2 ---
Zh->Ja 16.4 --- 6.9 --- 21.8 ---

NTCIR-8 Results
Intrinsic (BLEU) Extrinsic
JE EJ BLEU MAP Recall@100
KYOTO 22.22 24.29 17.25 0.1909 0.5258
Moses 29.08 35.27 24.01 0.1943 0.5701

Translation Samples
Input: 以下の添付図面に示す本発明に係るコネクタの保持構造の最良の形態について説明する。
Output: A description will be hereinafter given of a best mode a holding structure of a connector according to the present invention with reference to the attached drawings.

Alignment Evaluation Results (Ja-En)
Pre. Rec. AER
Step 1 86.20 44.54 41.24
Step 2 – 1 84.87 49.20 37.48
Step 2 – 2 86.19 58.71 29.98
Step 2 – 3 85.43 63.36 27.05
Step 2 – 4 82.68 65.31 26.83
Step 2 – 5 76.83 66.36 28.64
intersection 90.98 45.53 39.34
grow-final-and 80.00 60.48 31.00
grow-diag-final-and 77.86 61.93 30.92

Language Knowledge Engineering Lab.
Kyoto University