Introduction

This paper describes the machine translation system employed by SAS Institute Inc in the 1st Workshop on Asian Translation. We participate in two subtasks in this year’s WAT: Chinese to Japanese; English to Japanese. The sentence structure of Japanese is different with that of English/Chinese. Japanese is typically a Subject-Object-Verb (SOV) language while Chinese and English are Subject-Verb-Object (SVO) languages, as illustrated in the following Figure.

The statistic machine translation between Japanese and the SVO language is particularly difficult because of the long distance difference of word orders. We propose a simple syntactic reordering approach to transform Chinese/English into SVO languages. In addition, we apply the segmentation tool in SAS® Text Miner to the corpus and obtain improvement of the translation.

Background

- translate phrases as units; "Standard Model" used by Google Translate
- Limited capacity for long distance reordering

Syntax-based models [Li et al, 2006]
- Improve the translation
- System is complex and time consuming during decoding

Syntactic reordering approaches
- Effectively improve the translation results

System Architecture

1. Chinese to Japanese reordering rules

1.1. VP-rule: VP (VV AS (XXX)) → VP (XXX) VV AS

to move the verb (VV) and the auxiliary word (AS) behind VV to the end of the verb phrase (VP)

(a) Original

(b) Reordered

2. English to Japanese reordering rules

2.1. PP-rule: PP (P (XXX)) → PP (XXX) P

to move P to the end of PP.

(a) Original

(b) Reordered

Syntactic Reordering Approaches

1. Effect of the segmentation tool of SAS® Text Miner

Baseline (phrase-based model provided by the organizer):

Japanese: Human segmentation tool
Chinese: Stanford Word Segmenter
SAS segmentation tool of SAS® Text Miner for Chinese and Japanese

2. Chinese to Japanese translation

Baseline (phrase-based model provided by the organizer)

We gain 2.07 in BLEU scores compared with the baseline.

3. English to Japanese translation

Baseline (phrase-based model provided by the organizer)

We gain 3.13 in BLEU scores compared with the baseline.

Conclusion & Future Work

Conclusion:
- Introduce the system architecture of SAS at WAT 2014;
  - Describe the reordering approaches in detail;
  - Show experiments results to illustrate the effect of our system.

Future work:
- Consider Japanese Case Marker in the translation;
- Add more reordering rules on Chinese to Japanese translation;
- Attend the work to English to Chinese translation.