

12

Better Japanese understanding helps better translation

~Pre-ordering machine translation by deep syntactic analysis~

Abstract

Word order differences between Japanese and other languages cause serious translation errors by machine translation. We propose a method that **reorders Japanese words** into English or Chinese word order prior to machine translation, to make the translation easy.

Our key idea: reordering Japanese *subject-object-verb* order into *subject-verb-object* order by **deep syntactic analysis**. Typed dependency analysis in Japanese to identify syntactic relations between modifier and modifiee words. These relations reorder Japanese words and more **easily and accurately** translate the reordered Japanese sentences into English/Chinese. Such machine translation **technologies will ease language barriers** in global communication.

<Japanese>

彼は長い尻尾の猫を見た

Our method
in last year

彼は 長い 尻尾の 猫を 見た
Subject Object Predicate

Conventional
Machine
Translation

Re-ordering by
searching over
possible
permutations

Problem① Untyped dependency
(w/ post-process to identify Subj/Obj)
Problem② Chunk-based analysis
(complex word-level reordering)

He is a long tail of cat saw

Error in word order

New
method

は 彼 見た を 猫 の 長い 尻尾
Subject Predicate Object

① Typed dependency analysis

彼 は 長い 尻尾 の 猫 を 見た
Subject Noun-modifier Noun-modifier Object Auxiliary

② Pre-ordering

は 彼 見た 猫 の 長い 尻尾
Subject Auxiliary Object Noun-modifier Noun-modifier

③ Translation

He saw a cat with a long tail

Correct word order

Key Technology①: Dependency analysis

- Identify syntactic relations between words [1]
(conventional methods in Japanese are chunk-based)
- Accurate analysis same as English & Chinese

Key Technology②: Pre-ordering

- From Japanese SOV to English/Chinese SVO
- Arrangement for adjectives, particles, & auxiliary verbs

Key Technology③: Statistical MT [2]

- Learn translation patterns from translation data
- Easily solvable with (almost) monotone translation

Related works

- [1] T. Tanaka and M. Nagata, "A Study on Typed Dependency Analysis in Japanese," in Proc. ANLP, 2015. (In Japanese)
- [2] M. Nagata, K. Sudoh, J. Suzuki, Y. Akiba, T. Hirao, H. Tsukada, "Recent Innovations in NTT's Statistical Machine Translation," NTT Technical Review, Vol. 11, No. 12, 2013.

Contact

Katsuhito Sudoh Linguistic Intelligence Research Group, Innovative Information Laboratory
E-mail : sudoh.katsuhito(at)lab.ntt.co.jp

