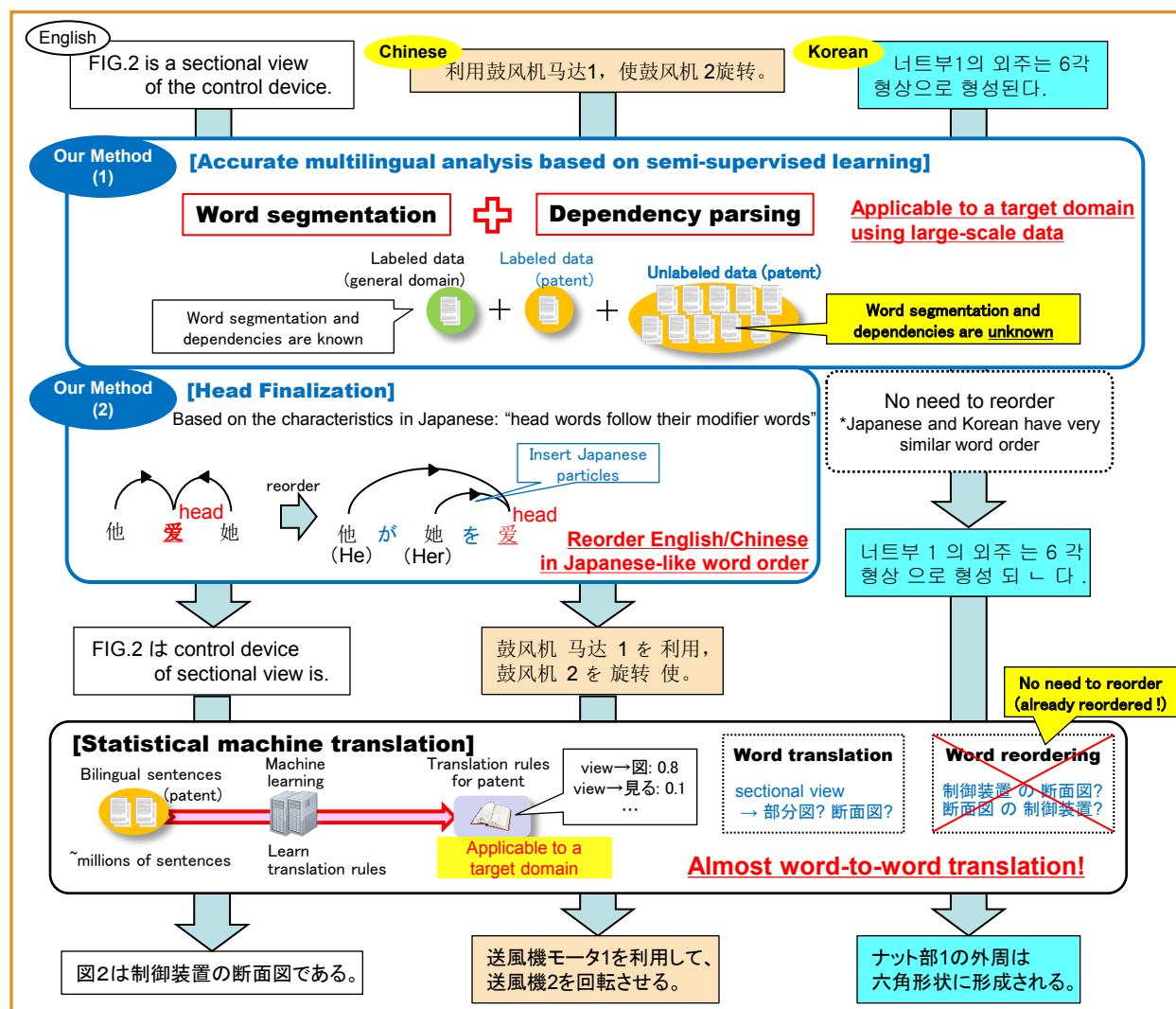




How can we overcome language barriers?

Statistical machine translation from foreign languages to Japanese

Abstract— Statistical machine translation is a promising machine translation technology that can learn “how to translate” automatically from bilingual sentence pairs. However, machine translation between Japanese and other languages is known to be difficult due to large differences in word order. Our technology overcomes this problem by (1) performing an accurate multilingual analysis based on a state of the art machine learning technique and (2) reordering foreign words into Japanese word order, based on the linguistic characteristics of Japanese. This enables accurate foreign-to-Japanese machine translation for written technical documents such as manuals and patents. In future, we plan to overcome language barriers for other kinds of documents and spoken languages.



Related works

- [1] J. Suzuki, H. Isozaki, X. Carreras, M. Collins, “An Empirical Study of Semi-supervised Structures Conditional Models for Dependency Parsing,” in *Proc. EMNLP*, pp. 551-560, 2009.
- [2] J. Suzuki, K. Duh, M. Nagata, “Simultaneous Natural Language Analysis based on Augmented Lagrangian,” in *Proc. Annual Meeting of ANLP*, pp. 1284-1287, 2012 (in Japanese).
- [3] H. Isozaki, K. Sudoh, H. Tsukada, K. Duh, “HPSG-based Preprocessing for English-to-Japanese Translation,” *ACM TALIP*, Vol. 11, 2012.

Contact

Katsuhito Sudoh Linguistic Intelligence Research Group, Innovative Communication Laboratory
E-mail : sudoh.katsuhito[at]lab.ntt.co.jp (Please replace {at} with @)