

言語の壁をどう越える？

～英語・中国語・韓国語から日本語への統計翻訳～

どんな研究

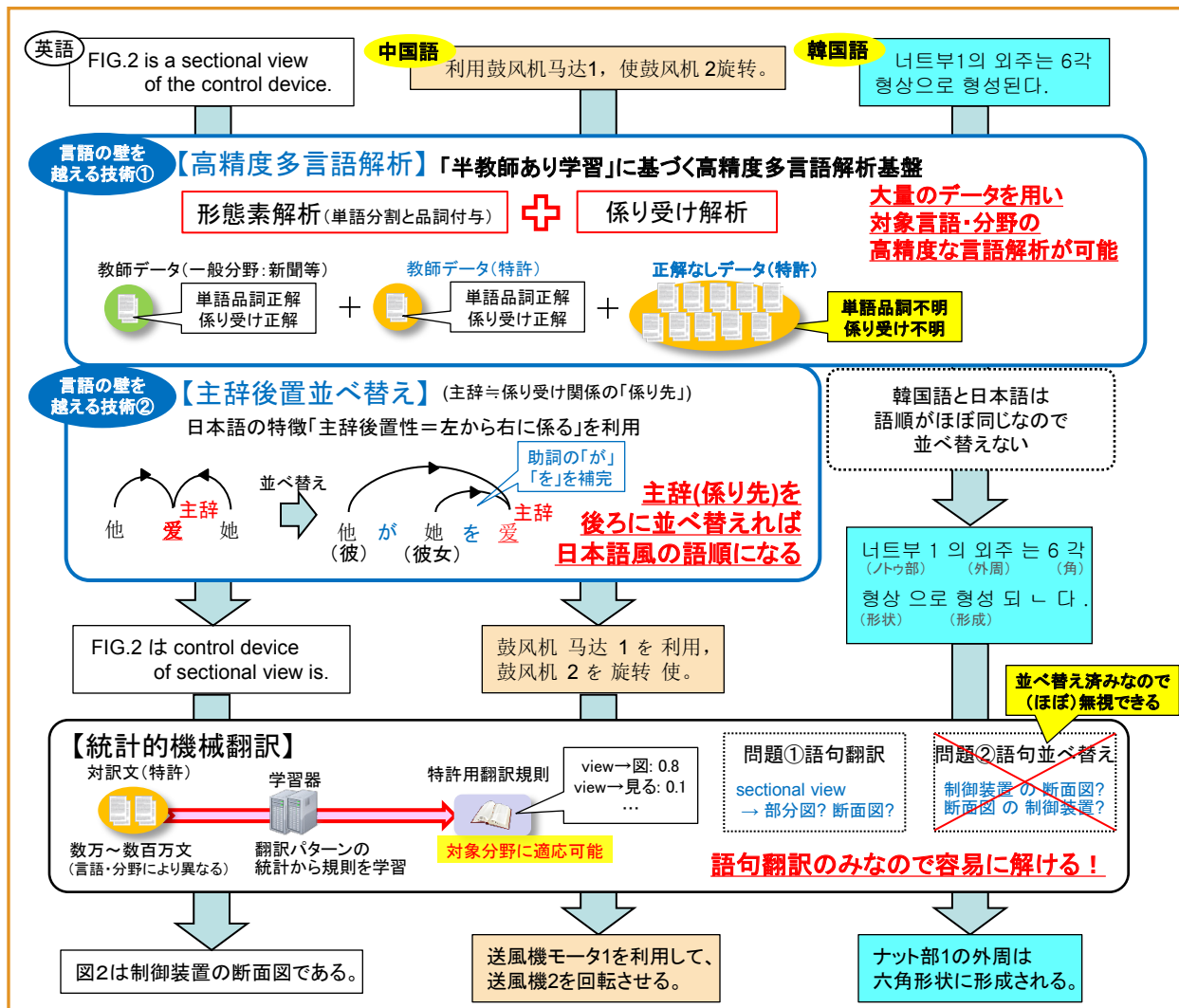
対訳例文を大量に使うことで翻訳の規則を確率的にモデル化し、機械翻訳を実現する「統計的機械翻訳」に関する研究です。この展示では、**高精度なことばの解析**と、**語順の違いを克服する並べ替え技術**とを組み合わせ、英語・中国語・韓国語を日本語に翻訳する技術を紹介します。

どこが凄い

多言語の学習データを独自に整備し、英語で培ってきた高度な言語解析技術を**多言語に拡張**しました。英語や中国語は日本語と大きく語順が異なりますが、この解析結果を用いて「翻訳前に」**日本語風に単語を並べ替える**ことで、高精度な日本語への翻訳を可能にします。

めざす未来

外国語で書かれた特許やマニュアル等の技術文書を日本語で探したり読んだりすることができるようになり、海外の情報を効率的に参照することができるようになります。将来的にはより一般的な文書や話しことばへも対象を広げ、言語の壁を越えたコミュニケーションを実現します。



関連文献

- [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] 鈴木潤, Kevin Duh, 永田昌明, "拡張ラグランジュ緩和を用いた同時自然言語解析法," 言語処理学会年次大会, pp.1284-1287, 2012.
- [3] H. Isozaki, K. Sudoh, H. Tsukada, K. Duh, "HPSG-based Preprocessing for English-to-Japanese Translation," *ACM TALIP*, Vol. 11, Issue 3, Article No. 8, 2012.

連絡先

須藤 克仁 (Katsuhito Sudoh) 協創情報研究部 言語知能研究グループ
E-mail : sudoh.katsuhito[at]lab.ntt.co.jp ({at}の部分>@に置き換えてください)