

日中英統計的機械翻訳のための事前並べ替え手法

Pre-ordering Methods for Chinese, English, and Japanese Statistical Machine Translation

韓丹 Dan Han, 星野翔 Sho Hoshino, 宮尾祐介 Yusuke Miyao

どんな研究？

- 日英翻訳・中日翻訳の精度が向上する並べ替え手法
- 中国語を日本語風、日本語を英語風にする規則
- 深い構文解析による言語の理解と翻訳

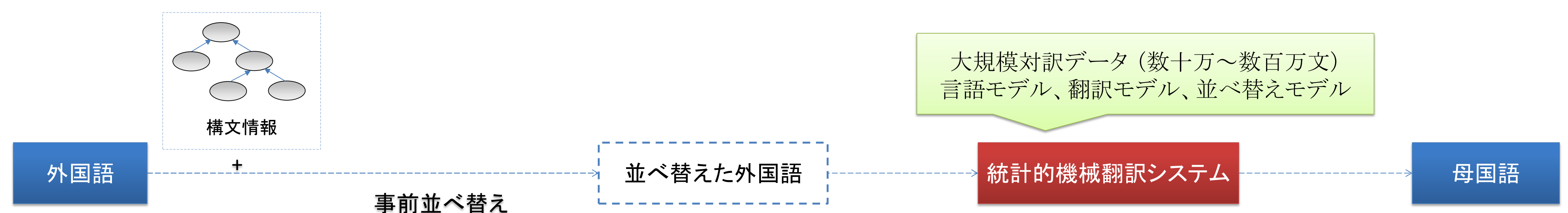
何がわかる？

- 翻訳／機械翻訳とは何か？研究上の難しさと問題点
- なぜ高度な機械翻訳が実用化されていないのか？
- 万能翻訳器(ほんやくコンニャク)は実現可能か？

機械による翻訳＝2つの言語の違いを理解して乗り越える

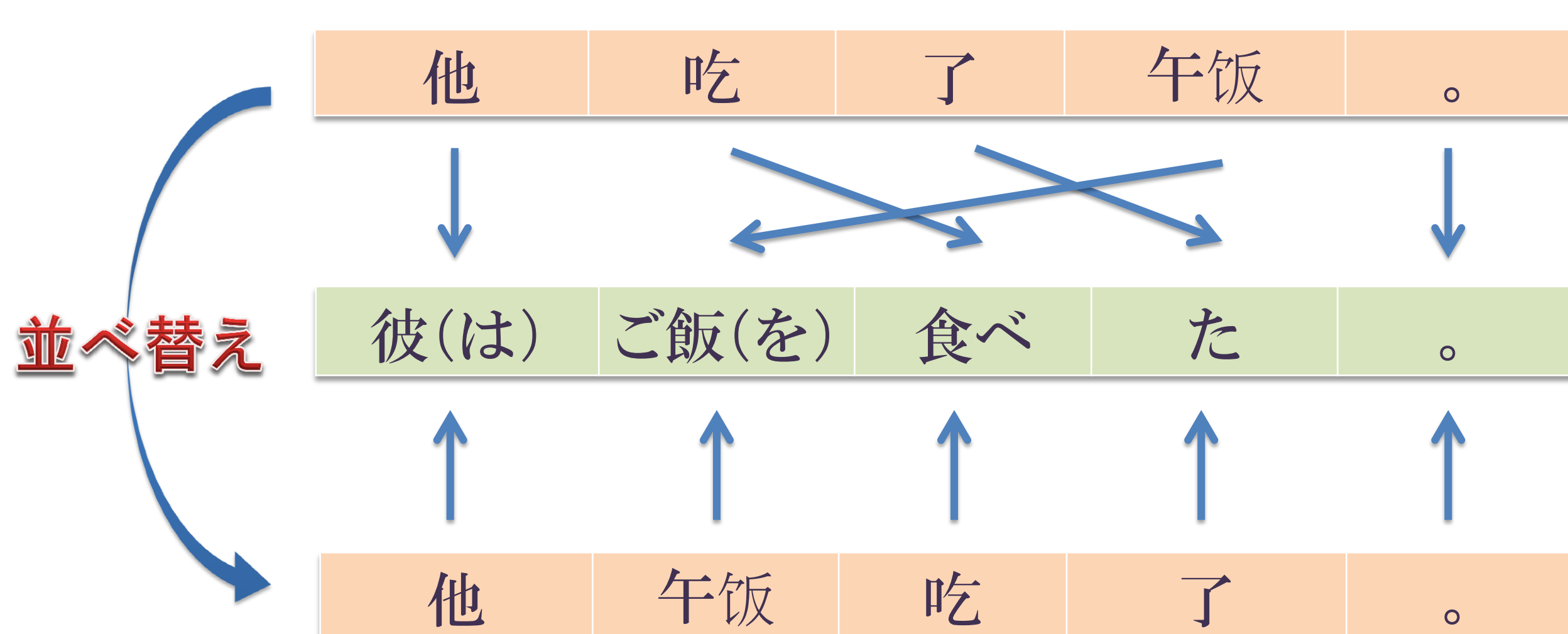


事前並べ替えによる 高精度翻訳

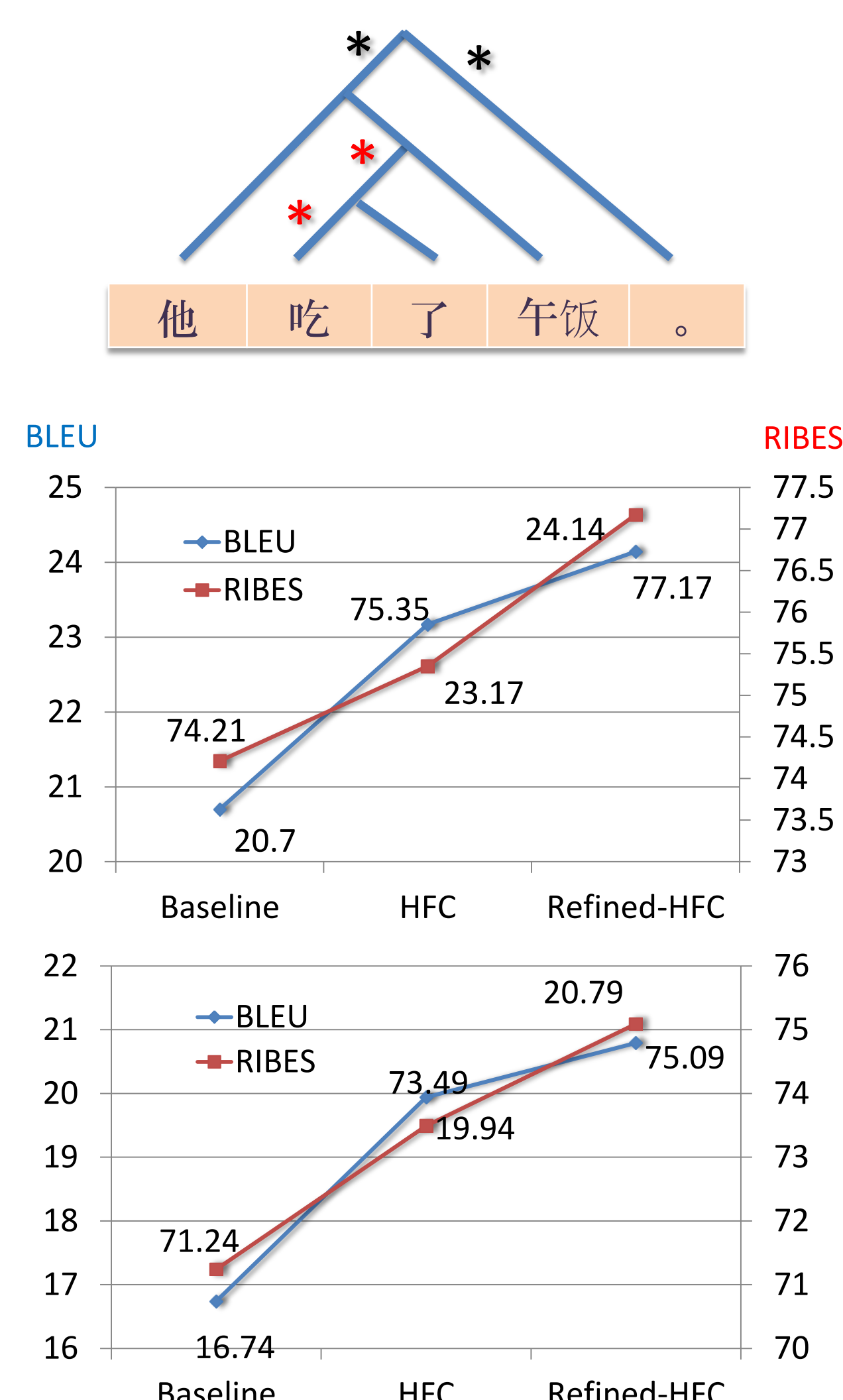


日本語と中国語、日本語と英語の違いは乗り越えられるか？

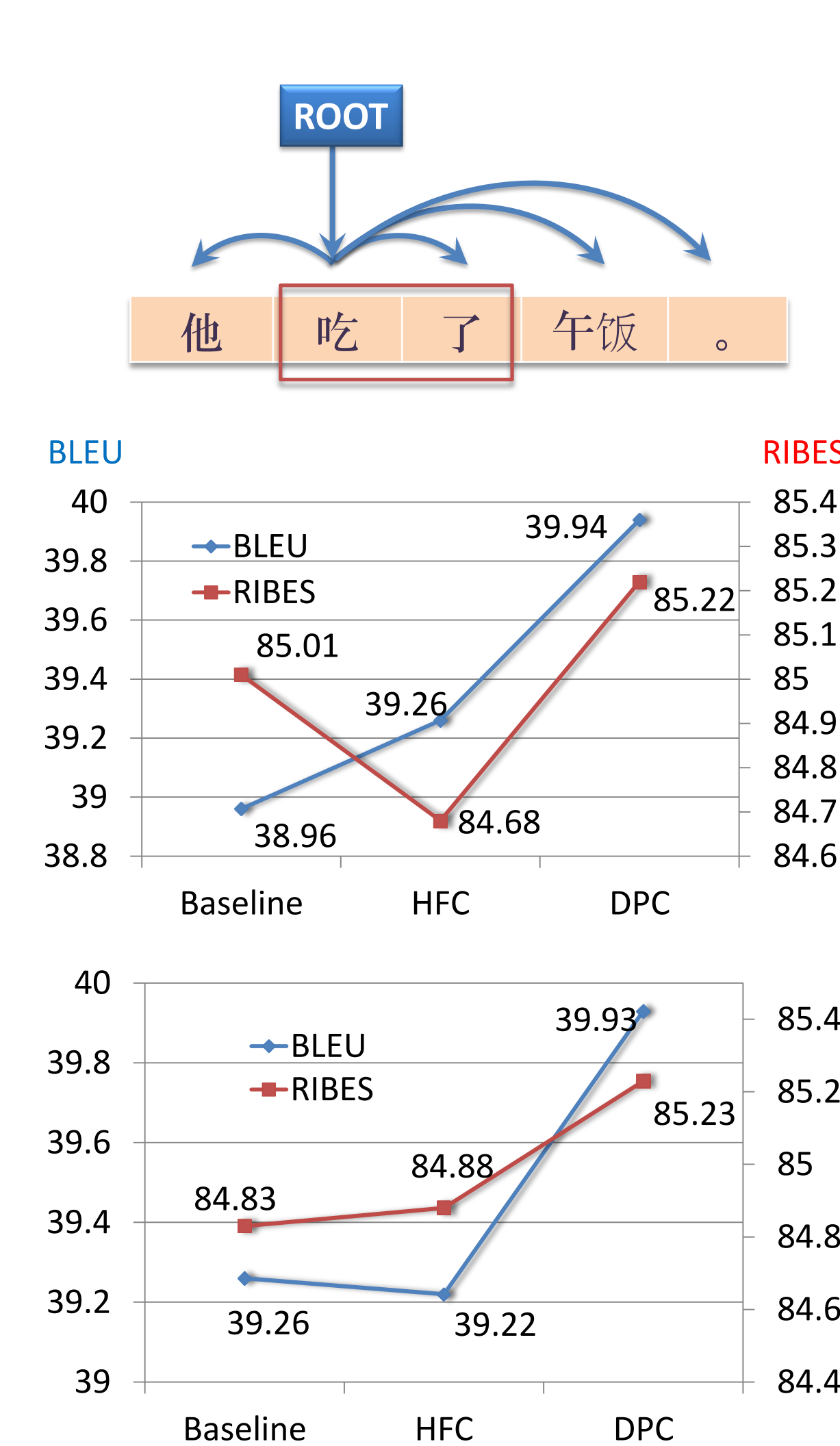
中国語 → 日本語



方法一: HPSG構文解析による(Refined-HFC)



方法二: 依存構文解析による(DPC)



日本語 → 英語 の克服: 2段階の事前並べ替え

- 文節間の並べ替え: SOV ⇒ SVO
- 文節内の並べ替え: 後置詞(助詞など) ⇒ 前置詞

私は京都へ行きます。

は私/行きます/へ京都/。

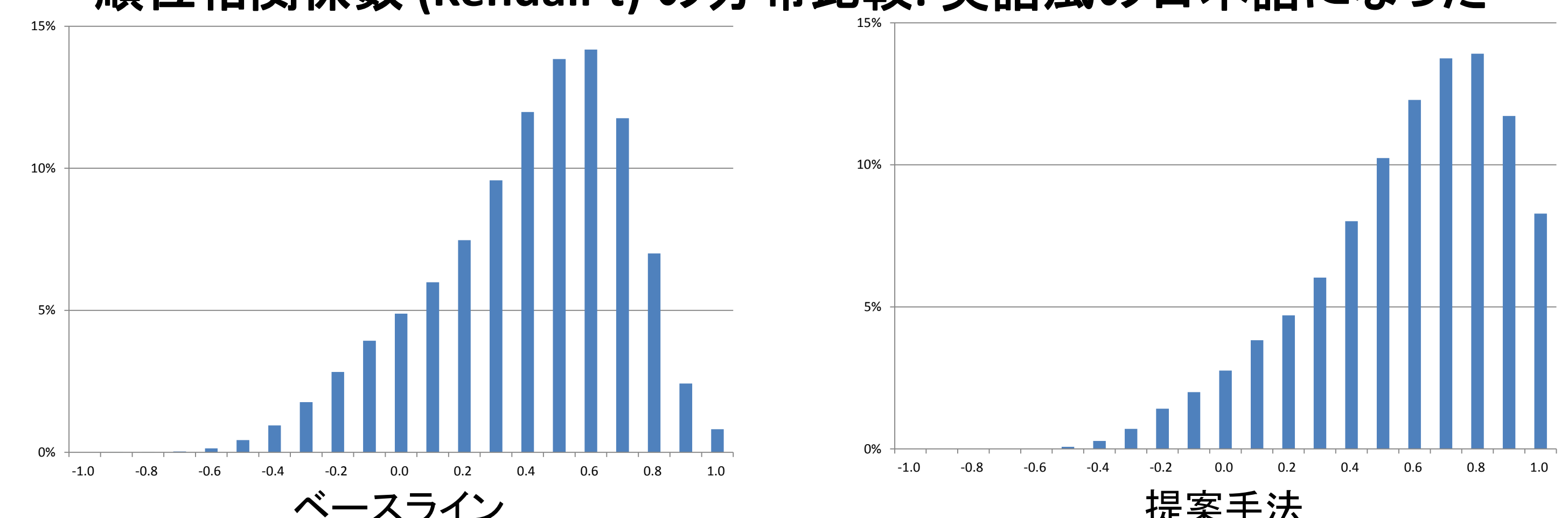
機械翻訳

I go to Kyoto.

320万文特許対訳データでの日英翻訳の比較

事前並べ替え手法	BLEU	RIBES
無し: ベースライン MeCabで単語分割	29.19	68.48
(Katz-Brown and Collins 2008)	27.59	66.10
(Komachi et al. 2006)	29.58	69.10
(Neubig et al. 2012) 1万文学習	29.93	70.15
提案手法 KNPで構文解析	30.65	72.26

順位相関係数 (Kendall's τ) の分布比較: 英語風の日本語になった



参考文献(詳細についてはこちらをご覧ください):

- Dan Han et al. (2012) Head finalization reordering for Chinese-to-Japanese machine translation. Proc. of SSST-6, pages 57-66.
- Dan Han et al. (2013) Using unlabeled dependency parsing for pre-reordering for Chinese-to-Japanese statistical machine translation. Proc. of HyTra2013.
- 星野ら (2013) 日英統計的機械翻訳のための述語項構造に基づく事前並べ替え, NLP2013.

連絡先: 宮尾祐介(Yusuke Miyao) / 国立情報学研究所 コンテンツ科学研究系 准教授

TEL/FAX : 03-4212-2590

Email : yusuke@nii.ac.jp