

TYPE OF INDUSTRY

科学技術・大学

情報通信研究機構

NICT 先端研究

(200)

機械翻訳(MT)の技術は、翻訳を起点言語のテキストを目標言語のテキストに変換する問題」とみなして研究されてきた。しかし実際には、テキストが表す内容が等価であるだけでは翻訳とはなりえ

ない。例えば、製品のマニュアルや特許の出願書類、企業のパレスリリースなどを翻訳する場合同様に、産業現場、対象分野の慣習、おいても、MTの出力、既存の用語の訳出方法、掲載媒体に応じて定められた句読法の仕度など、テキストの外側にある情報を適切に参照する必要がある。

MT技術の(テキスト変換の意味での)品質の向上に伴い、産業現場の修正という2種類の修正がなされる。(a)は文字の使い分けの必要性を示す例だが、他にも語句、句読法、記号の用法など、細かい仕様に従う必要がある。(b)のMTは、ニューズ記事が、他にも語句、句読法、記号の用法など、細かい仕様に従う必要がある。(c)のMTは、ローマ教皇に関する記事であったため、これらは

機械翻訳 参照すべき情報の分析

ユニバーサルコミュニケーション研究所
先進的音声翻訳研究開発推進センター 藤田 篤
先進的翻訳技術研究室 主任研究員

05年奈良先端大博士後期課程修了。国内の大学に研究員、教員として9年間勤務後、14年より現職。計算言語学、自然言語処理の研究に従事。特に、同義の言語表現(言い換え・翻訳)の認識・生成処理に関心を持つ。博士(工学)。

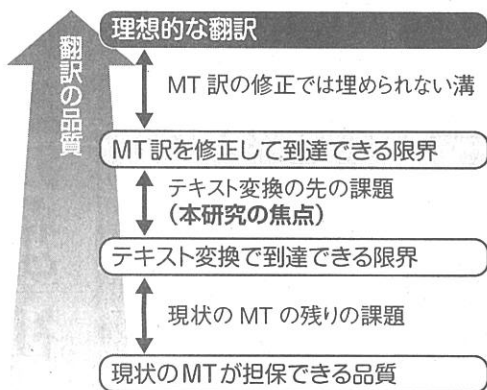


MTの修正事例

必要な情報の種類	原文中の表現	MT訳中の表現	修正訳中の表現
(a) 詳細なスタイル仕様	drawing	抽せん	抽選
(b) 既訳の用語集	John Paul	ジョン・ポール	ヨハネ・パウロ
(c) 分野固有の知識	3-0	3対0	3勝0敗
(d) 外部の関連文書	brother	兄弟	兄

既訳に修正された。(c)は米大リーグの

翻訳の品質と機械翻訳(MT)の課題



投手の戦績についての頼(できる)文書を入力言及であったため、点(できる)で初めてこの数ではなく勝敗が正し曖昧性を解消できた。(d)は起點文書 MTを真の翻訳に近い中の表現の曖昧性(兄づけるべく、私たちはか弟か)を解消できない。現在、このような情報かかった例である。修正を適切に参照する技術担当者、言及されて いる。(火曜日に掲載) いる人物に関する(信) いる。(火曜日に掲載)