

多段変換方式に基づく多言語翻訳手

Nur Suraya

松原 茂樹

村瀬 隆久

稲垣 康善

1 はじめに

近年、世界のボーダーレス化が進み、多言語間機械翻訳に対する需要が一段と高まりつつある。日本語と英語など、特定言語間の翻訳については、実用化レベルに近づきつつあるものの、それ以外の言語間翻訳については依然として今後の研究課題となっている。そこで本稿では、比較的研究が進んでいる言語間の翻訳技術を活用した多言語翻訳の一方式として、多段変換に基づく翻訳手法について述べる。本手法の応用例として、日本語からマレー語への機械翻訳の実現を試みる。

2 多段変換による多言語間翻訳

機械翻訳には、変換フェーズを介して実現するトランスファー方式と、中間言語とよばれる、ある共通の言語への翻訳によるインターリング方式の2つの方法がある。両者ともに一長一短があり、対象とする言語ごとに適した方式を使用することが望ましい。本研究では、トランスファー方式とインターリング方式を融合した方式に基づく多言語翻訳の実現を目指している。同族言語など、言語的距離が近い言語ごとにグループを構成し、その中の一つの言語をグループ内の中間言語と定める。グループ内の翻訳は、各言語とその中間言語との間の変換により、また、グループにまたがった言語間翻訳は、各言語とその中間言語、並びに、各グループの中間言語間など、多段階の変換によって、それぞれ実現する。図1左は、屈折言語、膠着言語のグループを構成し、それぞれ英語、日本語を中間言語とした場合の翻訳の構成例である。例えば、仏語 (French) から韓国語 (Korean) への翻訳は、(1) 仏-英、(2) 英-日、(3) 日-韓の3段階の変換によって実現する。このようなアプローチの利用可能性は従来から論じられているものの、実際の試みはこれまでほとんどない。

3 日本語-マレー語機械翻訳への応用

多段変換に基づく翻訳の一応用として、日本語-マレー語翻訳システムの実現を試みている。本システムは、図1左の構成に基づき、英語を介した2段階の翻訳、すなわち、日-英、及び英-マ翻訳から構成される。このうち、日本語から英語への翻訳には市販のソフトウェアを使用する。一方、英語とマレー語は、同じ屈折語に属するものの、いくつかの構文的な違いが存在するため、構文変換ルールを作成することにより英-マ翻訳を実現する(図1右)。すなわち、文脈自由文法を用いて英語文を解析し、作成した構文木に対して構文変換ルール及び対訳辞書を適用することにより、対応するマレー語文を生成する。

構文変換ルールを獲得するために、英語とマレー語の小規模なパラレルコーパスを作成し、その構文的違いについて調べた。その結果、

- 修飾・被修飾関係にある形容詞と普通名詞が、マレー語では、普通名詞、形容詞の順で生起し、名詞句を構成する。

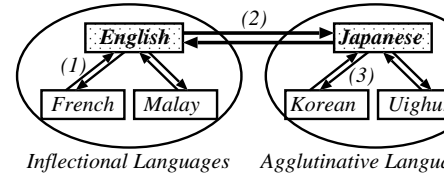


図1: 多段変換による多言語翻訳

表1: 英マ翻訳

英マ構文変換ルール	ルール
$JJ NN \Rightarrow NN JJ$	形容
$VBZ \Rightarrow \epsilon$	be動
$NNP NN \Rightarrow NN NNP$	固有

- 英語の be 動詞に相当する単語は、マレー語では生起し、文を構成する。

など、多くが品詞の並びに関する相違点があり、これらに基づき、本稿で述べたように構文変換ルールを作成した(表1)。本稿では、Penn Treebank で策定された基準に準拠する

Penn Treebank に収録されている日本語-マレー語対訳辞書を用いて英マ翻訳を行った。578文に付与された構文変換ルール、及び435項目の対訳辞書を適用することにより、文が正しく翻訳された。

4 おわりに

本稿では、多段変換に基づく多言語間翻訳システムの実現について述べた。日本語-マレー語機械翻訳について述べたが、日本語-マレー語機械翻訳について述べた。J/M が試作されているが[2]、これにより実現している。一方、日本語-マレー語機械翻訳について述べた。より、日英翻訳処理を修正することにより、日本語と同族である韓国語からマレー語への翻訳を実現している。

参考文献

[1] Marcus, M. et al.: Building a Large-Scale Morphologically Rich Treebank, *Computational Linguistics*

[2] Ogura, K., Bond, F. and Ooyama, T.: A Japanese-Malay Translation System, *Proceedings of the 1998 Conference on Computational Linguistics*