

機械翻訳システム構築ツールと その日本語-モンゴル語機械翻訳への適用

竹島 志起 黒木 泰行 サレンチモグ 谷口 誘一 松本 忠博

岐阜大学工学部応用情報学科

{takeshima,kuroki,saran,taniguchi,tad}@mat.info.gifu-u.ac.jp

1 はじめに

中国・内モンゴル自治区では書き言葉として、ウイグル系の伝統的なモンゴル文字（以下、モンゴル文字）で書かれる蒙古語文語 [1] が使われている。本稿におけるモンゴル語はこの書記言語を指す¹。モンゴル文字は計算機上での扱いが最も難しい文字と言われ²、1999年に Unicode 3.0 に追加された後も実装が伴わない状態が続いた³。モンゴル語は電子化とその標準化の遅れもあり、機械翻訳の対象としてはまだほとんど扱われていない（キリル文字による現代モンゴル語から日本語への機械翻訳については江原ら [2] の研究がある）。

我々は、日本語から多言語への翻訳を目的として開発されたパターン変換型機械翻訳エンジン jaw [3] と翻訳規則エディタ (jawEditor) から成る機械翻訳システム構築ツールの再構築・改良版の開発を行っており、これを用いて日本語からモンゴル語への機械翻訳を試みている。モンゴル文字の入出力については、OS に搭載されるようになった標準機能を利用した。本稿では、開発中の機械翻訳システム構築ツールと、その日本語-モンゴル語機械翻訳への適用について述べる。

2 機械翻訳システム構築ツール

2.1 機械翻訳エンジン jaw による処理の流れ

図 1 に jaw による翻訳処理の概略を示す。機械翻訳エンジン jaw は入力された日本語文を文節構造解析システム ibukiC [4] により解析し、その結果を日本語の構造を表す木構造 (Input Tree : IT) として保持する。次に IT を日本語パターンの辞書と照合し、IT がど

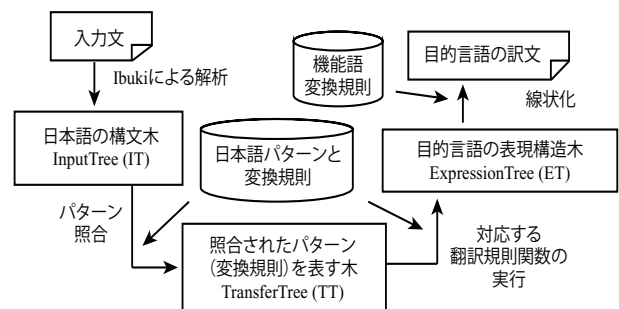


図 1: 機械翻訳エンジン jaw の処理の流れ

のような日本語表現パターンの組によって構成されているかということを示す木構造 (Transfer Tree : TT) を作成する。jaw は各日本語表現パターン毎に目的言語の表現構造への変換規則をプログラムコードとして持っており、この変換規則を用いて TT を目的言語の表現木構造 (Expression Tree : ET) に変換する。最後に ET の各ノードが持つ線状化プログラムを呼び出すことにより ET から目的言語テキストを生成する。

2.2 jawEditor

jaw において翻訳の際に照合する日本語表現パターンとそれに対応する変換規則は jawEditor という補助ツールを用いて登録する。jaw では目的言語の表現構造はクラスオブジェクトとして管理されており、日本語の表現構造を目的言語の表現構造へ変換するプログラムコードは jawEditor によって登録された規則から生成される。

3 日本語-モンゴル語機械翻訳への適用

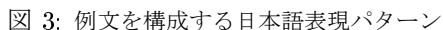
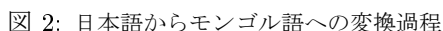
モンゴル語や日本語の文法書等から例文を取得して、対訳例文を作成し、日本語入力文から正しいモンゴル語文が得られるように翻訳規則を jaw に与えることで、日本語-モンゴル語機械翻訳を試みた。

以下、日本語文“私の母は毎朝花に水をあげます。”

¹モンゴル国ではキリル文字によるモンゴル語（ハルハ方言）が用いられており、小沢 [1] はこれを現代モンゴル語と称している。

²モンゴル文字は、アラビア文字と同じくアラム文字を起源とするが、縦書きで左から右に改行する。各文字は複数の形を持っており、単語内での位置などによって多いもので 8 通りに書き分けられる。

³Microsoft Windows では Vista からモンゴル文字のサポートが始まった。



3.1 例文を構成する日本語表現パターン

この日本語表現パターンは **jawEditor** で登録する。
図 5 は **jawEditor** で図 3 の Base パターン-1 を登録するときの画面、図 6 は図 3 の Base パターン-1 に対応する変換規則を登録するときの画面である。

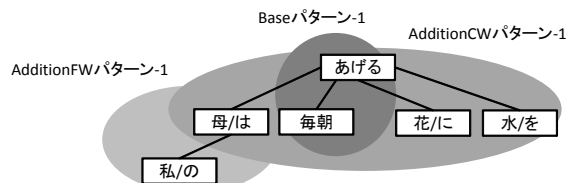


図 5: jawEditor における日本語表現パターン登録画面

図 6: jawEditor における変換登録画面

語順

Verb クラスは格要素，自分自身を修飾する副詞，状況語などをメンバとして持つ．jaw ではそれらをモンゴル語の語順に線状化していく．まずはじめに状況語を線状化し，次にモンゴル語においては，日本語と同じく動詞の格要素，副詞の語順は自由である．語順にも意味が含まれると考えられるので，それらの語順は日本語の語順に準じるようにした⁶．最後に動詞自体を翻訳し，句読点，疑問符などの文末の処理を行う．図 2 の例文の場合は，まず日本語の語順に準じて主格の^母，動詞を修飾する副詞の^毎 ^朝，対格の^花が線状化され，最後に動詞が語形変化し訳文に加えられる．

⁶従来の jaw のシステムでは入力文から翻訳規則の木構造 TT を作成した時点で語順情報が失われていたが，目的言語の表現構造まで語順を保持するように改良した．

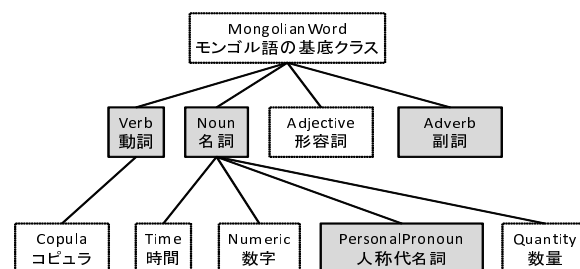


図 7: モンゴル語のクラス階層構造

3.3.2 Noun (名詞) クラス

名詞に付属する格助詞

名詞の後に格に応じた格助詞が付き名詞の格を表す．また同じ格を表す格助詞の形は 1 つではなく，名詞の語末に応じて形が変化する．以下の表に格助詞選択の規則の一部を示す．

表 2: モンゴル語の格助詞

属格(日本語の「の」に相当)	
語尾が母音で終わる場合	ᠤ
語尾が「ᠢ」以外の子音で終わる場合	ᠤ
語尾が子音「ᠢ」で終わる場合	ᠤ
対格(日本語の「を」に相当)	
語尾が母音で終わる場合	ᠤ
語尾が子音で終わる場合	ᠤ

主格に関しては格助詞は付かない．他には，与位格「に」，奪格「から」，造格「で」，共同格「と」等がある．例文の場合は名詞^母は動詞^母の主格となり，主格には格助詞が付かないので^母という形のままとなる．また名詞^花は対格となり，かつ語末が母音なので格助詞^ᠤが付与される⁷．

語順

Noun クラスは，名詞を修飾する形容詞，名詞と並列な名詞，名詞の属格となる名詞等のメンバを持つ．ま

⁷対格は語順，もしくは文脈によっては省略することができる．図 2 の例文の場合は，対格が主格よりも後にあるので格助詞を省略している．

3.3.3 PersonalPronoun (人称代名詞)

3.3.4 Adverb (副詞), Adjective (形容詞) クラス

4 翻訳結果と問題点

表 3: 翻訳例

翻訳例 1 は日本語の動詞が過去を表すときの例、翻訳例 2 はコピュラ文の例、翻訳例 3 は形容詞、副詞の例である。また、翻訳例 4 と翻訳例 5 は日本語においては同じあげるという動詞を用いているが、モンゴル語上ではそれぞれ違う表現となる例である。また解決が難しい問題として対格の格助詞を付けるかどうかの判別の問題がある。モンゴル語においては、対格を強調したり、対格の存在が一つに確定できる場合は必ず対格の格助詞を付けなくてはならない。翻訳例 6 と翻訳例 7 では日本語の語順が違う。翻訳例 7 では対格を強調するために対格の名詞（みかん）を文頭に持ってきている。これはシステム側で日本語の語順を保持することによって実装することができた。また「あの」「この」などのような指示詞が付く場合はそれに応じて判別できる。しかし文脈上、対格の名詞が強調または一つに特定されるものについては、jaw では一文ずつ翻訳を行っているために判別することができない。

5 おわりに

本稿ではパターン変換型機械翻訳システム構築ツールと、その日本語-モンゴル語機械翻訳システムへの適用について述べた。機械翻訳システム構築ツールは、現在公開を目指して整備を進めている。モンゴル語への機械翻訳については、より多くの例文を対象とした翻訳実験と言語的分析を通して問題点の発見と解決を行っていく予定である。

参考文献

- [1] 小沢重男, 蒙古語文語文法講義, 大学書林, 1997.
- [2] 江原暉将, 早田清冷, 木村展幸, 茶筌を用いたモンゴル語から日本語への機械翻訳, 言語処理学会第 11 回年次大会発表論文集, pp.534-537, 2005.
- [3] 池田尚志, “日本語からアジア諸言語への機械翻訳システムの構築奮闘記—中国語へ, ベトナム語へ, シンハラ語へ, 日本の手話へ—,” 日本語学 vol.28-12, no.10, pp62-70, 2009.
- [4] 池田尚志, 脇田貴之, 大口智也, “機能文節を導入した文節構造解析システム ibukiC (v0.20) について,” 言語処理学会第 14 回年次大会発表論文集, pp.221-224, 2008