

# 日本語から伝統的モンゴル語への機械翻訳システムの試作とその機能語の翻訳処理

竹嶋 志起      サレンチモグ      松本忠博

岐阜大学工学部応用情報学科

{takeshima,saran,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）で書写される蒙古語文語 [2] 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<sup>1</sup>。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<sup>2</sup>、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 [3] の研究がある）。モンゴル語と日本語は語順類型が同じ SOV 型であり、また共に膠着語である。そのことから単語置き換えで内容語部分についてはある程度まで翻訳できると考えられる。しかし動詞の語尾の変化によりテンス・アスペクト・モダリティ、副詞的な意味が表される点、助詞の形が付属する内容語の末尾の文字によって変化する点等、単純な単語置き換えだけでは翻訳できない要素もある。そのため、本研究では単語置き換えとパターン変換を併用した日本語から伝統的モンゴル語への機械翻訳システムを試作した。本稿では試作した機械翻訳システムの概要と機能語の翻訳処理について述べる。

## 2 モンゴル語とモンゴル文字

### 2.1 モンゴル語とモンゴル文字の特徴

- 伝統的モンゴル語では文字は縦書きで左から右に表記され、単語は分かち書きされる。

<sup>1</sup>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 [2] はこれを現代モンゴル語と称している。

<sup>2</sup>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。

- 単語中では日本語の草書体のように文字が繋がって表記され、文字の現れる位置によって形が変化する（図 1）。

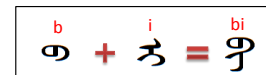


図 1: モンゴル語の文字の連結

- 母音調和

単語が男性母音 (a, o, u) か中性母音 (i) から構成される男性単語<sup>3</sup>と女性母音 (e, ö, ü) か中性母音から構成される女性単語に区別される。また、それによって動詞の語尾や付属する助詞の形が変化する<sup>4</sup>。

- 結合母音の法則

モンゴル文字を表記する際、単語内である文字に別の特定の文字が連結された場合、連結部分に特別な文字を付加しなければいけない場合がある。これは Visual Studio2010 の環境では、自動では付加されず文字入力の際に入力者が意図的に入力しなければならない。本システムにおいても動詞の語尾付与処理等、文字の結合処理をする場合がある。そのため、本システムでは以下の結合法則（図 2、図 3）に基づいて文字の結合処理関数を実装した。

#### ① 結合法則 1

男性（女性）単語の語末の A のいずれかの文字に B のいずれかの文字が結合するとき C の男性（女性）単語の文字が付加される（図 2）。

<sup>3</sup>中性母音からのみ構成される場合も男性単語に分類される。

<sup>4</sup>動詞の語尾や付属する助詞の形はそれが表現する意味によっても変化する。

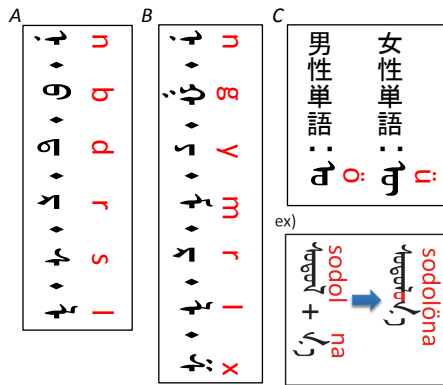


図 2: 結合法則 1

## ② 結合法則 2

図 3 中の A のいずれかの単語に B のいずれかの文字が結合するときに C の文字が付加される。

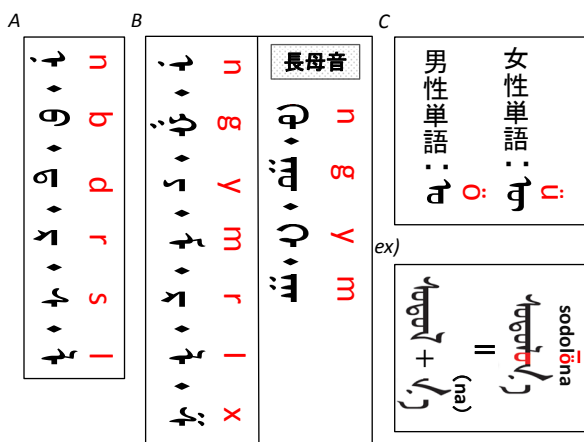


図 3: 結合法則 2

図は結合法則 2 の具体例である。

## ● キリル文字によるモンゴル語との違い

伝統的モンゴル語はモンゴル文字で縦書き表記されるという点以外にも、キリルモンゴル語では区別できない同音異義語を弱化石母音が表記されるため区別できるという点や、反対に伝統的モンゴル語には長母音を表す文字がない点などがキリルモンゴル語と異なる。つまりキリル文字によるモンゴル語と伝統的モンゴル語とは完全に対応しているわけではなく、翻字手法も研究されている [4] が、精度は 80.6 パーセントである。そのため本研究では日本語から伝統的モンゴル語へ直接翻訳している。

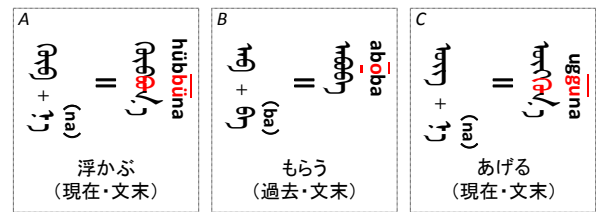


図 4: 結合法則 2 の具体例

## 3 システムの概要

### 3.1 ユーザインターフェイス

試作した翻訳システムのユーザインターフェイスを図 5 に示す。一般のテキスト表示コントロールは縦書きに対応していないので、本システムでは訳文の縦書き表示用に専用の表示部を作成した。

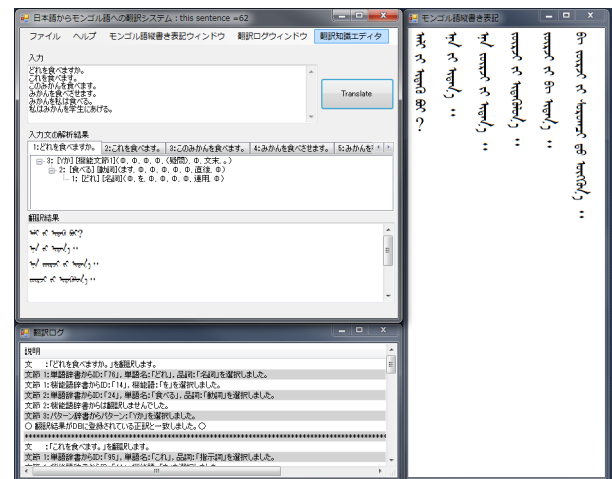


図 5: 翻訳システムのインターフェイス

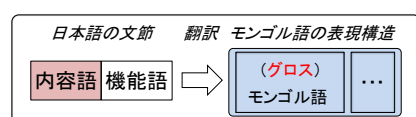
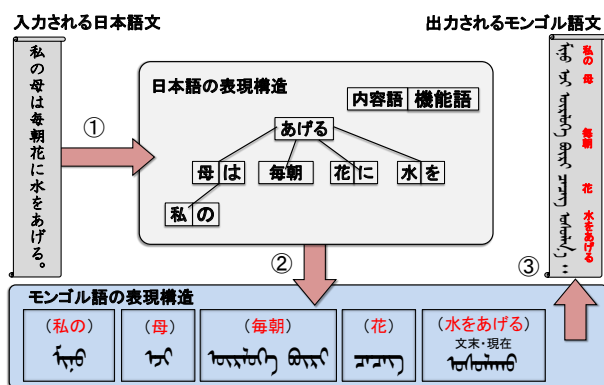
### 3.2 システムの処理の流れ

日本語文「私の母は毎朝花に水をあげる。」を本研究で試作したシステムで翻訳した場合の処理の流れを図 6 に示す。

- ① 入力文を日本語解析システム ibukiC[1] で形態素・係り受け解析をする。
- ② モンゴル語の言語知識辞書を参照し、日本語の文節毎にモンゴル語の表現構造に変換する。

日本語の解析結果からモンゴル語の表現構造への変換の過程を図 7 に示す。

- (1) 文節辞書の「私(の)」と照合



- (1) 私 の ⇒
- (2) 母 は ⇒
- (3) 毎朝 ⇒
- (4) 花 に ⇒
- (5) 水 を ⇒
- (6) あげる ⇒
- 図 7: モンゴル語の表現構造への変換の過程
- (2) 内容語辞書の「母」と照合
- (3) 内容語辞書の「毎朝」と照合
- (4) 内容語辞書の「花」と照合
- (5) 内容語辞書の「水」と照合
- (6) 日本語表現パターン「<水>(を)あげる」と照合
- ③ モンゴル語の表現構造を、語尾変化、助詞の付加等の処理を行いモンゴル語のテキストとして展開し出力する。

図6の例では、「水をあげる」にあたる動詞が基本形から、現在形・文末の語尾に変化している。

### 3.3 言語知識辞書

本システムでは翻訳に必要な伝統的モンゴル語の言語知識を内容語辞書，機能語辞書，文節辞書，パターン変換規則として扱う．図 8 に翻訳の際の照合の優先順位を示す．以下に各辞書について述べる．

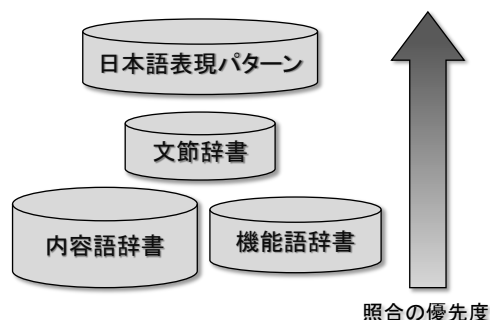


図 8: 試作したシステムの辞書

- 内容語辞書
- 伝統的モンゴル語の内容語を表す辞書．日本語の内容語とそれに対応するモンゴル語，意味属性，モンゴル語における品詞，読みを情報として持つ．
- 機能語辞書
- 伝統的モンゴル語の機能語を表す辞書．日本語の機能語とそれに対応するモンゴル語を情報として持つ．
- 文節辞書
- 日本語の内容語と機能語の組がモンゴル語の１つの単語に対応するものを扱う辞書．日本語の内容語，機能語とそれに対応するモンゴル語を情報として持つ．本システムでは入力文の解析結果を各文節の内容語，機能語の順に照合していくが，モンゴル語には<sup>私の</sup>あのように日本語の内容語と機能語の組み合わせを１つの単語で表すものもある．それらを翻訳するための辞書として本システムでは文節辞書を持つ．
- パターン変換規則
- 日本語の複数の文節に渡る表現を翻訳するためのパターン変換規則である「家を出る」の「家」は日本語においてはヲ格であるがモンゴル語においては奪格であるので格助詞もそれに応じたものにしなければならない．このようにモンゴル語と日本語で違う表現になるものを翻訳するために使用する．図 9 はパターン変換規則適用例である．

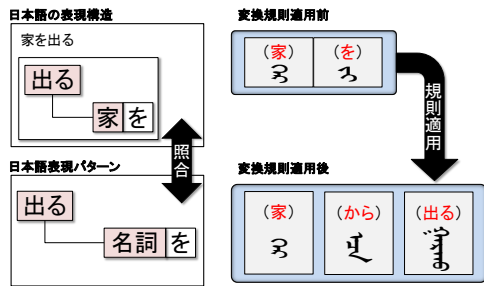


図 9: パターン変換規則適用例

## 4 機能語の翻訳処理

日本語の機能語部分はモンゴル語においては動詞の語尾変化や助詞として表れる．

### 4.1 動詞の語尾変化

伝統的モンゴル語では，動詞の時制や命令の種類，その出現位置によって動詞に語尾が付加される．本システムでは内容語を辞書と照合した後，機能語に応じて時制などの情報がモンゴル語の文節に付与される．訳文の出力のときにこの情報に応じて語尾を付加していく．本システムの単語辞書には動詞を形動詞形で登録しており，翻訳処理のときは登録されている単語から語幹のみを抽出する．次に文節に付与されている情報に応じて語尾を追加していく．付与される語尾には順序がある（表 1<sup>5</sup>）．

表 1: 動詞の語尾

意味	機能語	男性単語	女性単語	優先度
使役	させる	堂	里	1
受身	られる	通	通	2
連用	て	道	道	3
現在・文中		苗	苗	4
現在・文末		ふ	ふ	4
過去・文中	た	通	通	4
過去・文末	た	里	里	4
勧誘	う	つ	つ	4
依頼	てください	堂	里	4
命令	なさい	堂	里	4
強い命令	ろ	基本形	基本形	4

<sup>5</sup>表中の男性（女性）単語の列には動詞が男性（女性）単語のときに付加される語尾が，優先度には動詞に付加される順番が記されている．また表中の優先度が同一のものは同時には付加されない．

## 4.2 格助詞

表 2 は機能語とそれに応じたモンゴル語における格助詞である．

本システムでは機能語部分を翻訳するとき，格助詞に翻訳される場合はモンゴル語の表現構造の要素として，表 2 のどの格であるかを示すフラグを与える．そして出力時に付属する語の語末を参照し，文字形を選択する．基本的には機能語辞書の情報により表中の括弧内の機能語がその格として翻訳される．特定の条件下で表以外の機能語に翻訳する必要がある場合があるが，これはパターン変換規則を用いて翻訳している．

表 2: 格助詞の対応表

属格（日本語の「の」に相当）	奪格（日本語の「から」に相当）
付属する語の語末が母音の場合	通
付属する語の語末が「ハ」以外の子音の場合	通
付属する語の語末が子音「ハ」の場合	里
与位格（日本語の「に，へ」に相当）	共同格（日本語の「と」に相当）
付属する語の語末が母音か子音の「ハ，ハ，ハ，ハ」の場合	付属する語の語末が母音の場合
付属する語の語末が子音の「ハ，ハ，ハ，ハ，ハ」の場合	通
対格（日本語の「を」に相当）	
付属する語の語末が母音の場合	ろ
付属する語の語末が子音の場合	つ

## 5 おわりに

本稿では，日本語から伝統的モンゴル語への機械翻訳システムの試作とその機能語翻訳処理について述べた．今回の研究では，モンゴル語の言語知識として内容語辞書 215 項目，文節辞書 15 項目，機能語辞書 15 項目を登録した．今後は言語知識の登録数を増やしていく必要がある．

## 参考文献

- [1] 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
- [2] 小沢重男, 蒙古語文法講義, 大学書林, 1997.
- [3] 江原暉将, 早田清冷, 木村展幸, 茶釜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.
- [4] 満都拉, 藤井 敦, 石川 徹也 著伝統的モンゴル語と現代モンゴル語を対象とした双方向的な翻字手法情報処理学会論文誌 47(8), 2733–2745, 2006-08-15
- [5] フフバートル 著, 小沢重男 監修 「モンゴル語基礎文法」
- [6] 小沢重夫 著, 「モンゴル語四週間」
- [7] 東外大言語モジュール モンゴル語 URL[<http://www.coelang.tufs.ac.jp/modules/mn/-index.html>]