

# 茶笏を用いたモンゴル語の形態素解析

江原暉将

早田清冷

木村展幸

諏訪東京理科大学

東京外国語大学

(株)漢字情報サービス

[eharate@rs.suwa.tus.ac.jp](mailto:eharate@rs.suwa.tus.ac.jp)

[saksaha@spn1.speednet.ne.jp](mailto:saksaha@spn1.speednet.ne.jp)

[kimura@kiss.co.jp](mailto:kimura@kiss.co.jp)

## 1 はじめに

モンゴル語は日本語と語順が類似している<sup>(1)</sup>。語順空間内で見ると、モンゴル語は、朝鮮語、タミール語などとともに日本語にきわめて近い位置にくる(図1)。

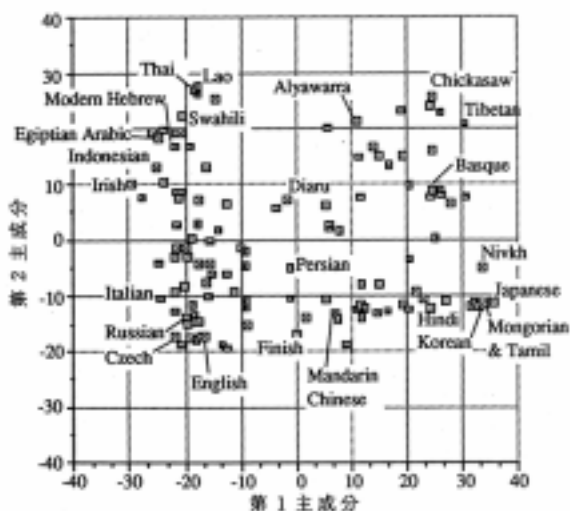


図1 語順空間での諸言語の配置<sup>(1)</sup>

また、モンゴル語は、名詞に格語尾(日本語の格助詞に相当)が付属したり、動詞に各種接辞が付属するなど日本語と類似した性質を持っている。

本文では、このようなモンゴル語と日本語の類似点に着目し、茶笏<sup>(2)</sup>を処理系として利用することでモンゴル語の形態素解析を行う試みについて述べる。

## 2 文字コード

モンゴル語の表記はキリル文字系によるもの(モンゴル新字)と伝統的モンゴル文字によるものの2種類がある。一方、茶笏は元来日本語を対象にして研究開発された形態素解析システムであり、使用できる文字コードも現時点では日本語の範囲に限られている。そこで、我々はJISコードに含まれるキリル文字を用いてモンゴル語を表記することとする。しかし、モンゴル新字にはJISコードのキリル文字セット

には含まれていない2種の文字が存在する。θとΥである(小文字はθとυ)。我々は、前者に対してギリシャ文字のΘを用い、後者に対しては英文字のVを用いることで対処した(小文字はθとv)。こうすることでモンゴル新字をJISコードの範囲内で扱うことができる。

## 3 モンゴル語の素描

文献<sup>(3)</sup>を参考にして、モンゴル語の概要を説明する。モンゴル語はアルタイ諸語の一つであり、膠着性を持つ。たとえば、улс(国)という名詞語幹に属格語尾ынをつけてулсын(国の)としたり、яв(行く)という動詞語幹に現在時を示す語尾наをつけてявнаという現在形を作ったりする。この点は、日本語と類似している。また、1章に示したように語順は日本語と良く似ており、平叙文中で、主語と目的語が動詞の前に来るいわゆるSOVの語順である。

日本語と異なる性質の一つとして、母音調和の存在がある。モンゴル語の母音調和は、語の第一音節に現れる母音の種類によって、第二音節以下に現れる母音の種類が制約を受けるという性質である<sup>注1</sup>。たとえば、яв(行く)の第一音節の母音яは男性母音であるので、現在時を示す語尾としてはнаが用いられるが、эхэл(始まる)の第一音節母音эは女性母音であるので、現在時を示す動詞語尾としては、наではなくнэが用いられ、эхэлнэとなる。ここで「語」とは、接辞などが付加してもよく、モンゴル新字ではスペースで区切られて表記される単位である。母音調和の規則を表1に示す。

形態素解析を行う上で考慮しなければならないモンゴル新字を用いた正書法の性質として、接合母音の存在がある。モンゴル語の子音は2種類に分けられる。共鳴子音(м, н, г, л, б,

<sup>1</sup> 本論文における母音調和の説明は、あくまでもモンゴル新字による正書法上の規則の説明であって実際の音声とは異なる。

в, р の七子音) と非共鳴子音である。共鳴子音は、語に現れるとき、その前後いずれか明瞭に聞かれるところに必ず母音を伴って書かれるという性質がある。例えば *в з э л* (見地、主義) という名詞では *л* が共鳴子音であるからその前に母音 *э* を伴う。しかし、*в з э л* に属格語尾 *ийн* が付加するときは、*в з э лийн* ではなく、*в з лийн* となる。*л* はその前後いずれかに母音を伴えば良いので前と後の双方に母音を持つ必要はなく、*и* が後にあるので *э* は書かれない。このように接合母音は存在したりしなかったりする性質がある。そのため、名詞語幹といえども形が変化する。

表 1 母音調和

性の名称	第一音節に現れる母音	第二音節以下に現れる母音
男性	, , , , ( )	, , , ,
男性	, , , ,	, , , ,
女性	, , , , ( )	, , , ,
女性	, , , , ( )	, , , ,
中性	, , , ,	, , , ,

#### 4 品詞分類・活用型・活用形・接続規則

茶筌でモンゴル語を解析させるには、形態品詞分類ファイル(**grammar.cha**)、活用型ファイル(**ctypes.cha**)、活用形ファイル(**cforms.cha**)、接続規則ファイル(**connect.cha**)をモンゴル語用に設定しなければならない。前節で述べた母音調和と接合母音という性質から、語形の活用(変化)が大きい<sup>2</sup>。これに品詞分類、活用型、活用形を用いることで対応する。これらの表の詳細については煩雑になるので付録に示すが、動詞、名詞を活用語としている<sup>3</sup>。内容語・機能語にかかわらず、母音調和による「性」を品詞細分類として扱った。

#### 5 辞書

自立語辞書については、解析実験が行えるだ

<sup>2</sup> ここでは「活用」など茶筌の用語をそのまま用いたため、言語学の用語とは異なる場合がある。

<sup>3</sup> 冊子型辞書と同様に、動詞は形動詞形未完了を見出し語とし、名詞は語幹を見出し語とする。

けの、ごく小規模のものしか作成していない。機能語辞書は文献(3)の巻末にある語尾・接辞索引を参考に作成した。付録に示す理由により、語尾は非活用語とした。茶筌辞書の意味情報の部分に日本語の訳語や語尾・接辞の意味などを記述した。解析結果から意味情報を取り出すことで簡易な蒙日機械翻訳が可能である。

#### 6 解析実験

次の2つの例文で解析実験を行った。

(1) *хааны зургийг зурна*  
 (2) *ахынхаа ирснийг мэдсэнгвй*  
 \_はスペースを意味している。解析結果を以下に示す。出力形式は「見出し(活用形)|見出し(基本形)|品詞 品詞細分類1 品詞細分類2|活用型|活用形|意味情報」である。

(1)  
*хаан* | *хаан* | 名詞 自立 男性 a | 名詞・共鳴子音 Ан | 基本形 | 王  
*ы* | *ы* | 名詞語尾 || | 属格  
 \_ | \_ | 空白 || |  
*зург* | *зураг* | 名詞 自立 男性 a | 名詞・共鳴子音 а г | 活用形 1 | 絵  
*ийг* | *ийг* | 名詞語尾 || | 対格  
 \_ | \_ | 空白 || |  
*зур* | *зурах* | 動詞 自立 男性 a | 動詞・v 共鳴子音 р а | 活用形 1 | 描く  
*на* | *на* | 動詞語尾 男性 a || | 終止形現在  
 EOS  
 (2)  
*ах* | *ах* | 名詞 自立 男性 a | 名詞・非共鳴子音 т з ц х а | 基本形 | 兄  
*ын* | *ын* | 名詞語尾 || | 属格  
*хаа* | *хаа* | 名詞語尾 || | 再帰人称  
 \_ | \_ | 空白 || |  
*ир* | *ирэх* | 動詞 自立 女性 э | 動詞・v 共鳴子音 р э | 活用形 1 | 来る  
*сн* | *сн* | 動詞語尾 女性 э || | 形動詞形過去  
*ийг* | *ийг* | 名詞語尾 || | 対格  
 \_ | \_ | 空白 || |  
*мэд* | *мэдэх* | 動詞 自立 女性 э | 動詞・非共鳴子音 д э | 活用形 1 | 知る  
*сэн* | *сэн* | 動詞語尾 女性 э || | 形動詞形過去  
*гвй* | *гвй* | 実詞語尾 女性 э || | 否定辞  
 EOS

意味情報を取り出すことで

- (1) 王\_属格 絵\_対格 描く\_終止形現在
- (2) 兄\_属格\_再帰人称 来る\_形動詞形過去\_対格 知る\_形動詞形過去\_否定辞

を得る。つまり、この例文は

- (1) 王の絵を描く
- (2) 自分の兄の来たのを知らなかったを意味している。

#### 7 今後の課題

本システムの研究開発は始めたばかりであり以下のように多くの課題がある。

・内容語辞書

実用的な形態素解析システムとするためには大規模な内容語辞書の整備が不可欠である。文献(4)に示す5万余語の辞書や文献(5)の中に単語リストがあるが、使用許可が得られるかどうかは不明である。

・モンゴル語コーパス

現状では、形態素コストや接続コストを人手で作成している。これらの値を統計的に定めるには大規模なモンゴル語コーパスが必要となる。

・蒙日対訳コーパス

用例ベースや統計ベースの機械翻訳を行うには、蒙日対訳コーパスが必要である。Voice of Mongolia で日本語放送を行っているようであり、このようなコーパスの存在の可能性がある。

・処理系

今回、形態素処理系として茶釜を用いたが、茶釜が利用できる文字系は限られており、ユニコードなど多言語を表記する文字系に対処していない。また、活用型の総数にも制約がある。処理系を独自開発するには多大な労力が必要であるため、茶釜をアジア諸言語の標準的な形態素解析処理系に発展させることが望まれる。そのためには、これらの点の改良が必要であろう。

## 8 おわりに

茶釜を処理系として利用したモンゴル語形態素解析の試みについて述べた。まだ、多くの課題が残されているが特に内容語辞書の整備が大きい。アジア諸言語の電子化辞書として CICC のプロジェクトで研究開発されたものがあり、広範な利用が望まれるが、その中にはモンゴル語の電子化辞書は含まれていない。いずれにしても実用的なモンゴル語解析システムや日蒙・蒙日機械翻訳システムなどを早期に実現し、モンゴル（内モンゴル、外モンゴルなど）と日本の間のコミュニケーションの充実に寄与したい。

### 参考文献

- (1) 江原暉将：多次元尺度構成法を用いた語順パラメータの関係付け、言語処理学会第 1 回年次大会発表論文集、pp.173-176, Mar., 1995.
- (2) 松本裕治ほか：形態素解析システム「茶釜」version 2.2.1 使用説明書、Dec., 2000.  
<http://chasen.aist-nara.ac.jp/chasen/doc/chasen-2.2.1-j.pdf>
- (3) 小沢重男：モンゴール語四週間、大学書林、1986.
- (4) 小沢重男：現代モンゴル語辞典、大学書林、1994.
- (5) 中里致元：ほんとに楽しいモンゴル語、Oct., 1999,  
[http://texa.human.is.tohoku.ac.jp/~chigen/cgn\\_idx.htm](http://texa.human.is.tohoku.ac.jp/~chigen/cgn_idx.htm)

## 付録 品詞・活用型・活用形・接続規則

・内容語品詞分類

内容語の品詞分類を表 2 に示す。品詞細分類 1 は自立と接辞があり、内容語としては、自立である。品詞細分類 2 は、表 1 に示す 5 種類の性種別がある。表 2 に示すほかに、代名詞、数詞、接続詞、間投詞、記号類などがあるが、現時点では辞書に記述がないため内容語品詞に加えていない。今後、内容語の品詞は増える可能性がある。

表 2 内容語品詞

品詞大分類	品詞細分類1	品詞細分類2
動詞	自立	性
名詞	自立	性
形容詞	自立	性
副詞		性
未知語		
空白		

・機能語品詞分類

機能語の品詞分類を表 3 に示す。動詞・名詞・形容詞の接辞は、内容語に後接して当該の品詞を作る接辞である。例えば、дуу（歌）という名詞にлаという動詞形成接辞が後接しさらに形動詞形未完了語尾xが接続して、дуулах（歌う）という動詞を作る。

表 3 機能語品詞

品詞大分類	品詞細分類1	品詞細分類2
動詞	接辞	性
名詞	接辞	性
形容詞	接辞	性
動詞語尾		性
名詞語尾		性
実詞語尾		性
複数接辞		性

一方、語尾にはна（現在時を示す動詞語尾）やийн（属格を示す名詞語尾）などがある。

・活用型・活用形

母音調和や接合母音に対応するために、語尾が付加できる動詞、名詞（自立、接辞とも）は活用品詞として取り扱う。語尾も母音調和によって付加する内容語に制約があるため、活用品詞とするのが本来であろうが、茶釜の活用型数の上限（256）があるために活用品詞とはしなかった。例えば、現在時を示す動詞語尾は基本形としてнаを見出し語として立て、на、но、нэ、нθ、нэ（それぞれ、男性а、男性о、女性э、女性θ、中性）と活用させる

