茶筌を用いたモンゴル語から日本語への機械翻訳

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　　saksaha@spn1.speednet.ne.jp　　kimura@kiss.co.jp

1　はじめに

モンゴル語と日本語とは語順が類似していることを利用して、日本語形態素解析システム「茶筌(1)」を用いた蒙日機械翻訳の研究を行っている。本文では文献(2)の続報として、その後の展開について述べる。

モンゴル語の処理を扱った文献(3)、(4)では形態素解析を table look up で行っているが、我々の方法はラティス構造を用いるものである。

2　品詞体系

モンゴル語の形態素解析のための品詞体系は、母音調和による連接制約を記述するために複雑になる。茶筌の品詞分類が持つ階層構造を利用して、品詞体系を構築した。品詞体系は 5 階層になっている。最上位レベルの品詞体系を表 1 に示す[1]。

表 1　第1レベルの品詞

動詞	動詞語尾
動詞不変	名詞語尾
名詞	未知語
名詞不変	空白
形容詞	終止符
副詞	コンマ
接続詞	開括弧
間投詞	閉括弧
後置詞	数字

最上位レベルの品詞は、日本語とほぼ同じである。日本語の格助詞に相当する語は格語尾として名詞語尾に含まれる。また、助動詞は第 3 レベルの品詞で接辞として扱われている。大きな違いは、動詞と名詞に 2 種類あることである。動詞と動詞不変の差は、接合母音の有無によって語幹が変化する動詞が単なる動詞であり、接合母音が無く語幹が変化しない動詞が動詞不変である。名詞と名詞不変の差も同じである。第 1 レベルの品詞体系は文献(5)、(6)を参考にして構成した。また、記号などについては、まだ不十分な部分がある。

第 2 レベルの品詞は表 2 に示す語の性別である。モンゴル語の単語には男性と女性と中性の性別があり、男性と女性は、さらに 2 種類に細分される。単語の性別により、接辞や語尾の連接が制約される。

表 2　第2レベルの品詞

男性 a
男性 o
女性 э
女性 θ
中性

第 3 レベルの品詞は表 3 に示すように「自立」と「接辞」の別である。接辞に対しては、－記号の左の品詞に接続して右の品詞になるもので細分類されている。例えば「名詞－動詞接辞」は名詞に接続して動詞化する接辞である。

表 3　第3レベルの品詞

自立	副詞－動詞接辞
動詞－動詞接辞	動詞－名詞接辞
名詞－動詞接辞	名詞－名詞接辞
形容詞－動詞接辞	

第 4 レベルの品詞は語幹末の文字種別であり、表 4 に示す種別がある。

表 4　第4レベルの品詞

母音語幹
共鳴子音語幹
隠れ子音語幹
非共鳴子音語幹
ж ч ш語幹
ь 語幹

第 5 レベルの品詞は語幹末の文字種別の細分類であり、表 5 に示す種別がある。語幹末の文字種別に第 4 レベルと第 5 レベルの 2 つのレベ

[1] 品詞体系などは文献(2)と比較して大きく変更した。

ルを設けたのは、連接規則のデータ量を減らすためである。第 5 レベルの品詞は動詞不変と名詞不変の品詞にのみ適用される。動詞と名詞に対しては、同様の情報が活用型によって表現されている。

表 5　第5レベルの品詞

母音語幹	ц語幹
м語幹	х語幹
н語幹	ф語幹
г語幹	д語幹
л語幹	с語幹
б語幹	к語幹
в語幹	п語幹
р語幹	ж語幹
隠れｎ語幹	ш語幹
隠れｇ語幹	ч語幹
т語幹	ь語幹
з語幹	

3　活用型・活用形

モンゴル語は母音調和と接合母音の有無によって語形変化が大きい。この現象を記述するために活用型と活用形を用いる[2]。本文では、活用を語幹の変化のみに限っている。文献(2)では、動詞に対する活用が語尾を含めて記述されていたのと異なる。それによって、動詞と名詞の活用を統一して扱え、活用型数を減らすことができた。その代わり、後述する内容語辞書の動詞見出し語として、通常の冊子型辞書で採用している形動詞形未完了を用いることができず、語幹を用いている。例えば、図 2 の「ｖｖｃ」は冊子型辞書では「ｖｖｃｅｘ」が見出し語となるのが通例である。

活用型・活用形の一覧を付録 1 に示す。活用型は語幹末の音(文字)によって決まる。活用形は､主として接合母音の有無による変化である。接合母音としては母音調和の制約を満たすものを用いなければならない。

4　連接規則

連接規則は、文献(5)を参考にして作成した。連接制約は、品詞、活用型、活用形、語形自身によって決まる。接続コストは、手作業によって付与している。現時点での連接規則数は 6,207 である。連接規則の例を図 1 に示す。

[2] 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。

```
((( ((動詞 男性ａ 自立)))
 (((動詞 男性ａ 動詞－動詞接辞)))) 300)
((( ((動詞 男性ａ * 母音語幹) * 基本形))
 (((動詞語尾 男性ａ) * * ｎａ))) 200)
((( ((名詞不変 中性 動詞－名詞接辞) * * ш))
 (((動詞語尾 中性) * * ｇｖй))) 100)
((( ((名詞 男性ａ * 共鳴子音語幹) 共鳴子音ａｎ 活用形１))
 (((名詞語尾 男性ａ) * * ы))) 100)
```

図 1　連接規則の例

5　内容語辞書

辞書の項目は以下に示すとおりである。

- 基本形(辞書見出し語)
- 品詞(第 1 レベルから第 5 レベル)
- 活用型(動詞と名詞のみ)
- 日本語訳語(意味情報の項に記述)

内容語辞書の作成には、文献(7)の辞書を著者の許諾を得て使用した。また、文献(8)の単語リストも利用した。文献(8)のリストには日本語訳語がないため、訳語としては「？？」が振られている。内容語辞書の例を図 2 に示す。現時点での内容語辞書見出し数は 8,485 語である。

```
(品詞 (名詞 女性ｅ 自立 共鳴子音語幹)) ((見出し語 (ｖｚｅｇ 1000)) (活用型 共鳴子音ｅｇ) (意味情報 {ペン/ボールペン}) )

(品詞 (動詞 女性ｅ 自立 共鳴子音語幹)) ((見出し語 (ｖｇｖｖｌ 1100)) (活用型 共鳴子音ｌ) (意味情報 ？？) )

(品詞 (動詞不変 女性ｅ 自立 非共鳴子音語幹 ｃ語幹)) ((見出し語 (ｖｖｃ 1000)) (意味情報 {生まれる/生じる/成り立つ/発生する}) )
```

図 2　内容語辞書の例

6　機能語辞書

機能語辞書は文献(5)の巻末にある語尾・接辞索引を参考に作成した。機能語辞書の例を図 3 に示す。機能語辞書では、日本語訳語として、文法機能を記述し、典型的な訳語を【】で囲んで記述している。現時点での機能語辞書の見出し語数は 1,635 語である。

(品詞（名詞語尾　女性θ)) ((見出し語（и й г　100))
(意味情報　普通格変化対格【を】))

(品詞（名詞　女性θ　名詞－名詞接辞　共鳴子音語幹))
((見出し語（θ х θ н　500)) (活用型　共鳴子音θ н)
(意味情報　【に属する人、に属するもの】))

(品詞（動詞語尾　男性ｏ)) ((見出し語（ｏ х　100)) (意味情報　形動詞形現在と未来【するところの】))

図 3　機能語辞書の例

7　翻訳実験

小規模な翻訳実験を行った。といっても、現時点では、日本語生成部の作成が未了であるために、茶筌の解析結果から、日本語訳語のある意味情報のフィールドを切り出して日本語単語列(複数訳語を含む)として並べることしか出来ていない。表 6 に文献(5)の練習問題から引用した例文とそれを機械翻訳した結果および基準日本語訳文を示す。

8　おわりに

茶筌を処理系として利用したモンゴル語形態素解析と蒙日機械翻訳の試みについて述べた。今後の課題の一つとして、日本語生成部の構築がある。そのためには、意味情報の項に記述されている日本語訳語情報に加えて日本語品詞情報なども付加する必要がある。また、当然のことながら訳語選択の必要がある。さらにゼロ情報の補完も必要となる。例えば、モンゴル語では名詞語幹がそのままで主格や対格を表すことがある。これを日本語に訳す場合には、格助詞や係助詞の補完が必要となる(表 6 参照)。また、現在は数千語レベルである辞書の規模を拡大することも課題である。

参考文献

(1) 松本裕治ほか：形態素解析システム「茶筌」version 2.2.1 使用説明書、Dec., 2000.
(2) 江原暉将、早田清冷、木村展幸：茶筌を用いたモンゴル語の形態素解析，言語処理学会第 10 回年次大会発表論文集、pp.709-712, Mar., 2004.
(3) S. ENKHBAYAR ほか：音韻論的・形態論的制約を用いたモンゴル語句生成、情報処理学会研究報告，NL-162, pp.87-93, July, 2004.
(4) S. ENKHBAYAR ほか：音韻論的・形態論的制約を用いたモンゴル語形態素解析、情報処理学会研究報告，NL-164, pp.41-46, Nov., 2004.
(5) 小沢重男：モンゴール語四週間、大学書林,1986.
(6) 小沢重男：現代モンゴル語辞典、大学書林,1994.
(7) 清水幹夫：電子日蒙索引（辞典），2001.
http://mk-smz.hp.infoseek.co.jp//OnEJMD.shtml
(8) 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

表 6　蒙日翻訳例

モンゴル語原文	蒙日翻訳結果(日本語単語列)	基準日本語訳文
Нар　их гэрэлтэй байна.	{太陽/日}___{偉大な/偉い/おお/大きい/多く/盛んに/大/大変/とても/非常に/もの/余程}_ _{明かり/光線/光}_【と】___{在る/居	太陽は、非常に明るい光をもっている。
Гэрэл　их илчтэй　байна.	{明かり/光線/光}___{偉大な/偉い/おお/大きい/多く/盛んに/大/大変/とても/非常に/もの/余程}___熱_【と】___{在る/居る}_.	光は非常に高い熱をもっている。
Гэрлийн　илч ертθнцийн амьдралыг соёрхоно.	{明かり/光線/光}_【の】___熱___{宇宙/世界/世}_【の】___{一生/存在}_【を】___恵む_【する】_.	太陽の熱は、世界の生活を恵む。
Ѳвс, гэрлийн　х ѵчээр　ургана.	{草/干し草}_, _{明かり/光線/光}_【の】___{力/パワー/労力/腕力}_【で】___{繁る/生える/実る}_【する】_.	草は光のおがげで育つ。
Ѳвс　чийгийн　х ѵчээр　ургана.	{草/干し草}___{しっき/湿気/湿度}_【の】___{力/パワー/労力/腕力}_【で】___{繁る/生える/実る}_【する】_.	草は湿気の力で育つ。
Ѳвс　газрын шимээр ургана.	{草/干し草}___{地面/大地/土地/場所/陸上}_【の】___栄養_【で】___{繁る/生える/実る}_【する】_.	草は土地の栄養によって育つ。
Мал　θвсθθр амьдарна.	家畜___{草/干し草}_【で】___{生きる/暮らす/やっていく}_【する】_.	家畜は草によって生活する。
Монгол　хѵн, малаар амьдарна.	モンゴル___{人間/人/者}_, ___家畜_【で】___{生きる/暮らす/やっていく}_【する】_.	モンゴル人は家畜によって生活する。

付録１　活用型・活用形

活用型	基本形	活用形1	活用形2
隠れθｎ	*	ｎ	θｎ
隠れａｎ	*	ｎ	ａｎ
隠れｇ	*	ｇ	
隠れｉｎ	*	ｎ	ｉｎ
隠れｎ	*	ｎ	ｎ
隠れｏｎ	*	ｎ	ｏｎ
隠れｅｎ	*	ｎ	ｅｎ
共鳴子音vｂ	vｂ	ｂ	
共鳴子音vｖ	vｖ	ｖ	
共鳴子音vｇ	vｇ	ｇ	
共鳴子音vｌ	vｌ	ｌ	v
共鳴子音vｍ	vｍ	ｍ	
共鳴子音vｎ	vｎ	ｎ	v
共鳴子音vｒ	vｒ	ｒ	v
共鳴子音θｂ	θｂ	ｂ	
共鳴子音θｖ	θｖ	ｖ	
共鳴子音θｇ	θｇ	ｇ	
共鳴子音θｌ	θｌ	ｌ	θ
共鳴子音θｍ	θｍ	ｍ	
共鳴子音θｎ	θｎ	ｎ	θ
共鳴子音θｒ	θｒ	ｒ	θ
共鳴子音ａｂ	ａｂ	ｂ	
共鳴子音ａｖ	ａｖ	ｖ	
共鳴子音ａｇ	ａｇ	ｇ	
共鳴子音ａｌ	ａｌ	ｌ	ａ
共鳴子音ａｍ	ａｍ	ｍ	
共鳴子音ａｎ	ａｎ	ｎ	ａ
共鳴子音ａｒ	ａｒ	ｒ	ａ
共鳴子音ｉｂ	ｉｂ	ｂ	
共鳴子音ｉｖ	ｉｖ	ｖ	
共鳴子音ｉｇ	ｉｇ	ｇ	
共鳴子音ｉｌ	ｉｌ	ｌ	ｉ
共鳴子音ｉｍ	ｉｍ	ｍ	
共鳴子音ｉｎ	ｉｎ	ｎ	ｉ
共鳴子音ｉｒ	ｉｒ	ｒ	ｉ
共鳴子音ｌ	ｌ	ｌ	*
共鳴子音ｎ	ｎ	ｎ	*
共鳴子音ｏｂ	ｏｂ	ｂ	
共鳴子音ｏｖ	ｏｖ	ｖ	
共鳴子音ｏｇ	ｏｇ	ｇ	
共鳴子音ｏｌ	ｏｌ	ｌ	ｏ
共鳴子音ｏｍ	ｏｍ	ｍ	
共鳴子音ｏｎ	ｏｎ	ｎ	ｏ
共鳴子音ｏｒ	ｏｒ	ｒ	ｏ
共鳴子音ｒ	ｒ	ｒ	*

活用型	基本形	活用形1	活用形2
共鳴子音ｙｂ	ｙｂ	ｂ	
共鳴子音ｙｖ	ｙｖ	ｖ	
共鳴子音ｙｇ	ｙｇ	ｇ	
共鳴子音ｙｌ	ｙｌ	ｌ	ｙ
共鳴子音ｙｍ	ｙｍ	ｍ	
共鳴子音ｙｎ	ｙｎ	ｎ	ｙ
共鳴子音ｙｒ	ｙｒ	ｒ	ｙ
共鳴子音ｅｂ	ｅｂ	ｂ	
共鳴子音ｅｖ	ｅｖ	ｖ	
共鳴子音ｅｇ	ｅｇ	ｇ	
共鳴子音ｅｌ	ｅｌ	ｌ	ｅ
共鳴子音ｅｍ	ｅｍ	ｍ	
共鳴子音ｅｎ	ｅｎ	ｎ	ｅ
共鳴子音ｅｒ	ｅｒ	ｒ	ｅ
降り2重母音vｊ	vｊ	v	*
降り2重母音θｊ	θｊ	θ	*
降り2重母音ａｊ	ａｊ	ａ	*
降り2重母音ｅｊ	ｅｊ	ｅ	*
降り2重母音ёｊ	ёｊ	ё	*
降り2重母音ｏｊ	ｏｊ	ｏ	*
降り2重母音ｙｊ	ｙｊ	ｙ	*
降り2重母音ｅｊ	ｅｊ	ｅ	*
降り2重母音ｙｊ	ｙｊ	ｙ	*
子音ь	ь	*	ｉ
子音相当ｙ	ｙ	ｙ	
短母音v	v	*	
短母音θ	θ	*	
短母音ａ	ａ	*	
短母音ｉ	ｉ	*	
短母音ｏ	ｏ	*	
短母音ｙ	ｙ	*	
短母音ｅ	ｅ	*	
長母音ｉｊ	ｉｊ	ｉ	*
非共鳴子音vｄ	vｄ	ｄ	
非共鳴子音vｚ	vｚ	ｚ	
非共鳴子音vｚ	vｚ	ｚ	
非共鳴子音vｃ	vｃ	ｃ	
非共鳴子音vｔ	vｔ	ｔ	
非共鳴子音vｃ	vｃ	ｃ	
非共鳴子音vｃ	vｃ	ｃ	
非共鳴子音vｓ	vｓ	ｓ	
非共鳴子音θｄ	θｄ	ｄ	
非共鳴子音θｚ	θｚ	ｚ	
非共鳴子音θｚ	θｚ	ｚ	
非共鳴子音θｃ	θｃ	ｃ	

活用型	基本形	活用形1
非共鳴子音θｔ	θｔ	ｔ
非共鳴子音θｃ	θｃ	ｃ
非共鳴子音θｃ	θｃ	ｃ
非共鳴子音θｓ	θｓ	ｓ
非共鳴子音ａｄ	ａｄ	ｄ
非共鳴子音ａｚ	ａｚ	ｚ
非共鳴子音ａｚ	ａｚ	ｚ
非共鳴子音ａｃ	ａｃ	ｃ
非共鳴子音ａｔ	ａｔ	ｔ
非共鳴子音ａｃ	ａｃ	ｃ
非共鳴子音ａｃ	ａｃ	ｃ
非共鳴子音ａｓ	ａｓ	ｓ
非共鳴子音ｉｄ	ｉｄ	ｄ
非共鳴子音ｉｚ	ｉｚ	ｚ
非共鳴子音ｉｚ	ｉｚ	ｚ
非共鳴子音ｉｃ	ｉｃ	ｃ
非共鳴子音ｉｔ	ｉｔ	ｔ
非共鳴子音ｉｘ	ｉｘ	ｃ
非共鳴子音ｉｃ	ｉｃ	ｃ
非共鳴子音ｉｃ	ｉｃ	ｃ
非共鳴子音ｉｓ	ｉｓ	ｓ
非共鳴子音ｏｄ	ｏｄ	ｄ
非共鳴子音ｏｚ	ｏｚ	ｚ
非共鳴子音ｏｚ	ｏｚ	ｚ
非共鳴子音ｏｃ	ｏｃ	ｃ
非共鳴子音ｏｔ	ｏｔ	ｔ
非共鳴子音ｏｃ	ｏｃ	ｃ
非共鳴子音ｏｃ	ｏｃ	ｃ
非共鳴子音ｏｓ	ｏｓ	ｓ
非共鳴子音ｙｄ	ｙｄ	ｄ
非共鳴子音ｙｚ	ｙｚ	ｚ
非共鳴子音ｙｚ	ｙｚ	ｚ
非共鳴子音ｙｃ	ｙｃ	ｃ
非共鳴子音ｙｔ	ｙｔ	ｔ
非共鳴子音ｙｃ	ｙｃ	ｃ
非共鳴子音ｙｃ	ｙｃ	ｃ
非共鳴子音ｙｓ	ｙｓ	ｓ
非共鳴子音ｅｄ	ｅｄ	ｄ
非共鳴子音ｅｚ	ｅｚ	ｚ
非共鳴子音ｅｚ	ｅｚ	ｚ
非共鳴子音ｅｃ	ｅｃ	ｃ
非共鳴子音ｅｔ	ｅｔ	ｔ
非共鳴子音ｅｃ	ｅｃ	ｃ
非共鳴子音ｅｃ	ｅｃ	ｃ
非共鳴子音ｅｓ	ｅｓ	ｓ