

音韻論的・形態論的制約を用いたモンゴル語句生成

Sanduijav ENKHBAYAR[†] 宇津呂武仁[†] 佐藤 理史[†]

[†] 京都大学大学院 情報学研究科 知能情報学専攻
{enk, utsuro, sato}@pine.kuee.kyoto-u.ac.jp

本稿では、現時点で利用可能なモンゴル語の言語資源、特に、名詞・動詞の語幹のリスト、および、名詞・動詞に接続する語尾のリストから、モンゴル語の名詞句・動詞句を生成する手法を提案する。具体的には、名詞・動詞の語幹に語尾が接続する際の音韻論的・形態論的制約を整備し、語幹・語尾の語形変化の規則を作成する。評価実験の結果においては、8割以上の場合について、生成された名詞句・動詞句の中に正しい句候補が含まれるという性能が達成できた。

Mongolian Phrase Generation based on Phonological and Morphological Constraints

Sanduijav ENKHBAYAR[†], Takehito UTSURO[†], and Satoshi SATO[†]

[†] Dpt. Intelligence Sci. and Tech., Graduate School of Informatics, Kyoto University
{enk, utsuro, sato}@pine.kuee.kyoto-u.ac.jp

Using currently available Mongolian linguistic resources such as the lists of stems of nouns and verbs as well as the list of suffixes, this paper proposes a method of generating noun/verb phrases of the Mongolian language. More specifically, this paper examines phonological and morphological constraints on connecting stems of nouns/verbs and suffixes, and then, proposes inflection/conjugation rules for nouns/verbs. We experimentally show that, for more than 80% cases, correct noun/verb phrases can be found among the candidates of noun/verb phrases generated by the proposed method.

1. はじめに

モンゴル語においては、自立語の語幹に対して格を表す語尾や動詞の活用を表す語尾・接続助詞等が結合したものが句を構成し、ヨーロッパ言語と同様に、空白で区切られた句の列により文を構成する。ここで、モンゴル語の形態素解析の問題について考えると、この問題は、モンゴル語文中の名詞句や動詞句が与えられて、それらの句を名詞あるいは動詞の語幹と語尾とに分解することであると言える。この処理を実現するためには、名詞あるいは動詞の語幹に語尾が接続する際の接続可能性や語変形の規則性を明らかにする必要がある。また、例えば、他の言語からモンゴル語への機械翻訳などにおいては、名詞あるいは動詞の語幹および語尾が与えられると、その語幹・語尾の組に対する語変形や活用の過程を規則化し、名詞句あるいは動詞句を生成する機構を確立する必要がある。

ところが、現時点で利用可能なモンゴル語の言語資源としては、数千語程度の規模の単語について語幹情報が登録された電子辞書、および、ウェブ上で収集可能な新聞記事等の電子テキストが存在するにすぎない。また、モンゴル語に関して、名詞あるいは動詞の語幹と語尾の組から名詞句あるいは動詞句を生成するための言語知識や規則なども全く整備されていない。

そこで、本稿では、現時点で利用可能なモンゴル語の言語資源、特に、名詞・動詞の語幹のリスト、および、名詞・動詞に接続する語尾のリストから、モンゴル語の名詞句・動詞句を生成する手法を提案する。具体的には、名詞・動詞の語幹に語尾が接続する際の音韻論的・形態論的制約を整備し、語幹・語尾の語形変化の規則を作成する。評価実験の結果においては、8割以上の場合について、生成された名詞句・動詞句の中に正しい句候補が含まれるという性能が達成できた。

2. モンゴル語の文法

現代モンゴル語で使われる文字はキリル文字である。モンゴル語では、自立語の語幹に対して格を表す語尾や動詞の活用を表す語尾・接続助詞等が結合したものが句を構成し、ヨーロッパ言語と同様に、空白で区切られた句の列により文を構成する。モンゴル語の語順は日本語と同じSOVで、動詞が文末に位置し、その他の句の語順は比較的自由である。

通常、名詞の語幹には、数を表す語尾、格を表す語尾、再帰所属を表す語尾がこの順に接続する。名詞に接続する語尾の分類、および、各分類ごとの語尾の種類数を表1に示す。通常、同一の分類に対応する語尾には数種類の可能性があり、一つの名詞に接続する語尾を決定する際には、

表 1 名詞に接続する語尾の一覧

語尾の分類	語尾種類数
属格	7
对格	2
与位格	3
奪格	4
造格	4
共同格	3
再帰所属	4
合計	27

- (1) 名詞語幹
xγγxэд (子供)
- (2) 名詞語幹 + 複数語尾
xγγxдγγд (子供達)
- (3) 名詞語幹 + 複数語尾 + 格語尾
xγγxдγγдэд (子供達へ)
- (4) 名詞語幹 + 複数語尾 + 格語尾 + 所属語尾
xγγxдγγддээ (自分の子供達へ)

図 1 名詞の語形変化の例

その複数の可能性の中から，3. で述べる母音の接続制約，および，4.1 で述べる語幹・語尾の接続制約を満たす語尾が選ばれる．そして，4.2 の語形変化規則により，語幹・語尾が語形変化する．名詞の語幹にこれらの語尾が接続した場合の語形変化の例を図 1 に示す．

同様に，動詞の語幹に接続する語尾は，命令・願望類，叙述類，完了・習慣等を表す類，順序関係を表す類，等に分類される．動詞の活用語尾の分類，および，各分類ごとの語尾の種類数を表 2 に示す．動詞の場合も，同一の分類に対応する語尾には数種類の可能性があり，一つの動詞に接続する語尾を決定する際には，その複数の可能性の中から，3. で述べる母音の接続制約，および，4.1 で述べる語幹・語尾の接続制約を満たす語尾が選ばれる．そして，4.2 の語形変化規則により，語幹・語尾が語形変化する．動詞の語幹にこれらの語尾が接続して動詞が活用する例を図 2 に示す．

3. モンゴル語の母音字の接続制約

モンゴル語においては，名詞・動詞の語幹に語尾が接続する際に，名詞・動詞の語幹の末尾および語尾において語形変化・活用が起こる．本節では，この語尾の接続において，名詞・動詞に含まれる母音字と，語尾に含まれる母音字の間で，どのような接続制約が満たされる必要があるかについて述べる．なお，本節および次節で述べる内容は，文献 [1] に基づいており，日本語での用語等は文献 [3] に従っている．

3.1 モンゴル語の母音字と子音字

モンゴル語で用いる文字は全 35 文字で，母音字 13 字，子音字 20 字，記号文字 2 字から構成される．長母音，二重母音を含めたモンゴル語の母音の一覧を表 3 に，子音の一覧を表 4 に，それぞれ示す．母音字のうち補助母音

表 2 動詞の活用語尾の一覧

		活用の分類	語尾種類数
1	命令・願望	1 人称意思 1	3
2		1 人称意思 2	2
3		2 人称命令	0
4		2 人称勧告	4
5		2 人称催促	4
6		2 人称懇願	2
7		1-3 人称願望	4
8		1-3 人称懇念	2
9	叙述	現在・未来	4
10		単純過去	4
11		体験過去	4
12		伝聞過去	2
13	形動詞	完了	4
14		継続	4
15		予定	1
16		習慣	4
17		可能性	4
18	副動詞	連合	1
19		並列	2
20		分離	4
21		条件	8
22		継続	4
23		限界	4
24		即刻	4
25		随伴	4
26		付帯 1	2
27		付帯 2	4
合計			89

- (1) 動詞語幹
ud (食べる)
- (2) 動詞語幹 + 受動態語尾
udээд (食べられる)
- (3) 動詞語幹 + 使役態語尾
удүүл (食べさせる)
- (4) 動詞語幹 + 意志の語尾
удбе (食べよう)
- (5) 動詞語幹 + 単純過去の語尾
удэв (食べた)
- (6) 動詞語幹 + 完了の語尾
удчихсэн (食べてしまった)
- (7) 動詞語幹 + 従属節 (限界) 語尾
удтэж (食べるまで)
- (8) 動詞語幹 + 受動態語尾 + 単純過去の語尾
удээдэв (食べられた)

図 2 動詞の活用の例

字 1 は，多音字であり子音+母音とみなされる．また，補

表 5 モンゴル語の母音と性の関係

男性母音	<i>a, o, y, я, ё, ю, bi</i>
女性母音	<i>э, ө, ү, e, ю</i>
中性母音	<i>u</i>

表 3 モンゴル語の母音

基本母音字	<i>a, o, y, u, ə, θ, ɥ</i>
補助母音字 1 (子音 + 母音と見なされる)	<i>ɣ, e, ě, ɯ</i>
補助母音字 2 (長母音として扱う)	<i>bi</i>
補助母音字 3 (基本的に単独では使われない)	<i>ü</i>
長母音	<i>aa, oo, yy, uü, əə, θθ, ɥɥ</i>
二重母音	<i>aü, oü, yü, əü, ɥü</i>

表 4 モンゴル語の子音字

母音を必ず伴う子音字	<i>m, n, ɣ, ɣ, b, θ, p</i>
母音を伴わなくてもよい子音字	<i>ð, t, ʧ, ʃ, ʒ, c, u, x</i>
特殊子音字 (外来語に使われる)	<i>n, ʃ, u, κ</i>

表 6 モンゴル語の母音字の接続制約

語幹の母音字	語尾の母音
<i>a, y, ɣ</i>	<i>a, ɣ, aü</i>
<i>o, ě</i>	<i>o, ě, oü</i>
<i>ə, ɥ, u, e, ɯ</i>	<i>ə, e, əü</i>
<i>θ</i>	<i>θ</i>
<i>a, o, y, ɣ, ě, ɯ, bi</i>	<i>y, bi</i>
<i>ə, θ, ɥ, u, e, ɯ</i>	<i>ɥ, uü</i>

表 7 母音字の接続制約の例外: 語幹の末尾の文字と語尾の先頭字の接続制約

語幹の末尾の文字	語尾の先頭字
<i>ʃc, ʧ, u, u, ə, b</i>	<i>uü</i>

補助母音字 1 に基本母音を一つつけるという形でも使われる。補助母音字 2 は、単独で長母音として扱われる。補助母音字 3 は、単独で使われることはなく、他の母音とともに使われ二重母音を構成する。

また、モンゴル語の母音は、男性・女性・中性の三つの性を持ち、その内訳は表 5 となる。モンゴル語の単語の性は、強勢が置かれる母音の性によって決まる。

3.2 母音字の接続制約

動詞・名詞の語幹と語尾の接続の際に、双方の母音字の間に満たされるべき接続制約をまとめると、表 6 となる。この表は、語幹の母音字、および、それに接続可能な語尾の母音の組の一覧となっている^(注1)。なお、ここで接続可能な母音同士は、表 5 で示した性が同一のものとなっている。ただし、この場合、中性母音は女性母音として扱う。

また、表 6 の母音字の接続制約に対する例外として、名詞・動詞の語幹の末尾の文字と語尾の先頭字の間においては、表 7 に示す接続制約が満たされなければならない。ここで、表 7 に示す接続制約は、名詞・動詞の語幹の末尾

の文字が表 7 左側の文字となる場合で、しかも、語尾の候補として、表 7 右側の文字を先頭字として持つものが含まれる場合には、必ずその語尾が選ばなければならない、と解釈される。そして、このときには、表 6 の母音の接続制約は必ずしも満たされる必要はない。

4. モンゴル語の語幹・語尾の語形変化

通常、同一の分類に対応する語尾には数種類の可能性があり、一つの名詞あるいは動詞に接続する語尾を決定する際には、その複数の可能性の中から、まず、前節で述べた母音の接続制約を満たす語尾が選ばれ、さらに、4.1 で述べる語幹・語尾の接続制約を満たす語尾が選ばれる。そして、4.2 の語形変化規則により、語幹・語尾が語形変化する。

4.1 語幹・語尾接続制約

動詞・名詞の語幹の末尾と語尾の接続において満たされるべき接続制約をまとめると、表 8 となる。この表は、語幹の末尾、および、それに接続可能な語尾の組の一覧となっている。

4.2 語幹・語尾の語形変化

語幹・語尾の接続における語形変化の規則は、以下の四種類である。

- (1) 母音字消失の規則 (表 9)
- (2) 軟音符 *b* が母音字 *u* に変化する際の規則 (表 10)
- (3) つなぎの母音字の挿入規則 (表 11)
- (4) 母音以外のつなぎの文字の挿入規則 (表 12)

語幹・語尾の接続における語形変化の際に、語形変化後の語のアクセントが変化する場合がある。この場合、特に、アクセントが変化して、母音が発音されなくなることがあり、この発音されなくなった母音字が消失する。その際の規則性を記述したものが「母音字消失の規則」である。ただし、以下の場合には、必ずしも母音字が消失しなくてもよい。

- (1) 「母音を必ず伴う子音字」に伴っている母音が消失する場合
- (2) *n, ɣ* の直後の母音字
- (3) *ʃc, ʧ, u* 以外の子音字の直後の *u*
- (4) 固有名詞の母音字
- (5) 形動詞予定形の母音字

(注1): 2. で述べたように、同一の分類に対応する語尾には数種類の可能性があり、一つの名詞もしくは動詞に接続する語尾を決定する際には、その複数の可能性の中から、母音の接続制約、および、語幹・語尾の接続制約を満たす語尾が選ばれる。ここで、同一の分類に対応する語尾の候補同士が持ち得る母音の可能性は限られており、表 6 中の上半分もしくは下半分のどちらかの母音の組合せに限定される。したがって、語尾の分類が決まれば、その際に参照される母音字の接続制約は、表 6 中の上半分もしくは下半分のどちらかとなる。

表 8 語幹の末尾と語尾の接続制約

語幹の末尾	語尾
<i>n</i>	<i>z</i>
<i>c, x</i>	<i>m, ʉ</i>
母音	<i>a</i>
母音を必ず伴う子音字 + 軟音符 ₁	母音を伴わなくてもよい子音字
子音字 + 軟音符 _b	<i>ʃ, ɛ̃</i>
子音字 + 軟音符 _b	<i>zʉʉ</i>
母音	<i>e, ʃ, ɛ̃</i>

表 9 母音字消失の規則

	語の末尾	語尾の先頭	語形変化後
(i)	<i>u</i> 以外の母音	長母音	語の最後の母音が消失
(ii)	母音字 + 子音字	長母音	子音字 + 長母音
(iii)	母音字 ₁ + 子音字 ₁	子音字 ₂	子音字 ₁ + 母音字 ₂ + 子音字 ₂

表 10 軟音符₁が母音字 *i* に変化する際の規則

	語幹の末尾	語尾の先頭	語形変化後
(i)	子音字 + 軟音符 _b	長母音 (<i>xx</i>)	子音字 + <i>ux</i>
(ii)	子音字 + 軟音符 _b	母音を必ず伴う子音字	子音字 + <i>u</i> + 母音を必ず伴う子音字
(iii)	母音を伴わなくてもよい子音字 ₁ + 軟音符 _b	母音を伴わなくてもよい子音字 ₂	母音を伴わなくてもよい子音字 ₁ + <i>u</i> + 母音を伴わなくてもよい子音字 ₂
(iv)	子音字 + 軟音符 _b	<i>x</i> (形動詞予定形活用語尾)	子音字 + <i>ux</i>

表 11 つなぎの母音字の挿入規則

	語幹の末尾	語尾	語形変化後
(i)	母音を伴わなくてもよい子音字	子音字	母音を伴わなくてもよい子音字 + 母音字 + 子音字
(ii)	母音を必ず伴う子音字 ₁	母音を必ず伴う子音字 ₂	母音を必ず伴う子音字 ₁ + 母音字 + 母音を必ず伴う子音字 ₂
(iii)	<i>c, u</i>	<i>a</i>	<i>c, u + a</i> + 母音字

表 12 母音以外のつなぎの文字の挿入規則

	語幹の末尾	語尾の先頭	語形変化後
(i)	(女性語) 子音字	<i>e</i>	子音字 + 軟音符 _b + <i>e</i>
(ii)	(男性語) 子音字	<i>ʃ, ɛ̃</i>	子音字 + 硬音符 _b + <i>ʃ, ɛ̃</i>
(iii)	長母音 ₁	長母音 ₂	長母音 ₁ + <i>e</i> + 長母音 ₂
(iv)	長母音 ₁	長母音 ₂	長母音 ₁ + <i>ux</i> + 長母音 ₂

語幹・語尾の接続における語形変化の際に、軟音符_bが母音字 *u* に変換することがある。その際の規則性を記述したものが「軟音符_bが母音字 *u* に変化する際の規則」である。また、語幹・語尾の接続において子音が連続する場合は、つなぎの母音字を挿入する。その際の規則性を記述したものが「つなぎの母音字の挿入規則」である。その他の場合で、語幹・語尾の接続において、母音以外のつなぎの文字を挿入する場合もある。その際の規則性を記述したものが「母音以外のつなぎの文字の挿入規則」である。

5. モンゴル語句候補生成およびコーパスを利用した検証

5.1 手順

現時点で利用可能なモンゴル語の言語資源としては、数千語程度の規模の単語について語幹情報が登録された電子辞書、および、ウェブ上で収集可能な新聞記事等の電子

テキストが存在するにすぎない。そこで、本節では、それらの言語資源、名詞・動詞に接続する語尾のリスト、および、3. で述べた母音の接続制約、4. で述べた語幹・語尾の語形変化の規則を用いて、モンゴル語の名詞句・動詞句を生成する方法について述べる。ここでの句生成手順の概要は以下ようになる。

(1) 与えられた名詞もしくは動詞の語幹に対して、格や活用の分類に応じた語尾の全候補をまず求める。

(2) 3. で述べた母音の接続制約に基づいて、語尾の候補を絞り込む。

(3) 4. で述べた語幹・語尾の語形変化の規則を用いて、名詞・動詞の句候補を生成する。

(4) モンゴル語コーパスに対して、生成された句候補の検索を行ない、検索できなかった句候補を削除する。

通常、モンゴル語の名詞の語幹に接続する語尾の種類としては、複数語尾、主格(語尾なし)を含む7格語尾、再

表 13 名詞+属格の語形変化の例

語幹	<i>боловсрол</i>
語尾候補	<i>ны, ын, ы, ний, ий, ийн</i> (属格の全語尾候補) → <i>ны, ын, ы</i> (男性名詞に接続可能な語尾候補) → <i>ны, ын</i> (属格固有の接続制約: <i>ы</i> は語の末尾が <i>н</i> 以外には接続不可)
語形変化	(語幹+ <i>ны</i>) → <i>боловсролын</i> (語幹+ <i>ын</i>) → <i>боловсролын</i> (語幹+ <i>ын</i>) → <i>боловсролын</i> → (母音字消失規則 (ii)) → <i>боловсрын</i>
コーパスによる検証	<i>боловсролын</i>

表 14 名詞+与位格の語形変化の例

語幹	<i>боловсрол</i>
語尾候補	<i>д, т</i> (与位格の全語尾候補)
語形変化	(語幹+ <i>д</i>) → <i>боловсролд</i> (語幹+ <i>д</i>) → <i>боловсролд</i> → (与位格固有のつなぎの子音字 <i>н</i> の挿入) → <i>боловсролд</i> → (つなぎの母音字の挿入規則 (ii)) → <i>боловсрололд</i> (語幹+ <i>т</i>) → <i>боловсролт</i> (語幹+ <i>т</i>) → <i>боловсролт</i> → (与位格固有のつなぎの子音字 <i>н</i> の挿入) → <i>боловсролт</i> → (つなぎの母音字の挿入規則 (ii)) → <i>боловсролонт</i>
コーパスによる検証	<i>боловсролд</i>

表 15 動詞の活用「副動詞:並列」の例

語幹	<i>арилга</i>
語尾候補	<i>ж, ч</i> (副動詞:並列の全語尾候補)
活用	(語幹+ <i>ж</i>) → <i>арилгаж</i> (語幹+ <i>ч</i>) → <i>арилгач</i>
コーパスによる検証	<i>арилгаж</i>

表 16 動詞の活用「命令・願望:1-3 人称懸念」の例

語幹	<i>арилга</i>
語尾候補	<i>уузай, чцээй</i> (命令・願望:1-3 人称懸念の全語尾候補)
活用	(語幹+ <i>уузай</i>) → <i>арилгауузай</i> → (母音字消失規則 (i)) → <i>арилгуузай</i> (語幹+ <i>чцээй</i>) → <i>арилгаццээй</i> → (母音字消失規則 (i)) → <i>арилгццээй</i>
コーパスによる検証	<i>арилгуузай</i>

帰所属語尾がある。このうち、本稿では、複数語尾、主格を除いた7種類の語尾の分類(語尾の種類数では27種類)について、句候補生成手順を評価した。一方、動詞の活用については、27種類の活用の分類(語尾の種類数では89種類)について、句候補生成手順を評価した。

また、本稿で用いたモンゴル語コーパスは、ウェブ上のモンゴル語新聞一年半分のテキストを収集してコーパスとしたもの(延べ語数206万、異なり語数11万5千、30MBytes)である。

5.2 例

表13~16に、名詞+属格の語形変化の例、名詞+与位格の語形変化の例、動詞の活用「副動詞:並列」の例、および、動詞の活用「命令・願望:1-3 人称懸念」の例をそれぞれ示す。

表13の名詞+属格の語形変化の例においては、「*боловсрол*」(教育)という名詞に「~の」という意味の属格語尾が接続した場合の語形変化の様子を示す。まず、属格語尾の全候補として六種類の語尾が得られるが、このうち、男性名詞「*боловсрол*」に接続可能な語尾は三種類に絞られ

る。さらに、属格固有の語幹・語尾接続制約により、接続可能な語尾は二種類に絞られる。この二種類の語尾が語幹に接続すると、語形変化を伴わず語幹に語尾がそのまま接続した形の句候補が二種類、および、母音字消失規則(ii)が適用された形の句候補が一種類、合計三種類の句候補が生成される。そして、コーパスに対してこれらの句候補の検証を行なった結果、一種類の句候補のみが得られる。

表14の名詞+与位格の語形変化の例においては、「*боловсрол*」(教育)という名詞に「~に」という意味の与位格語尾が接続した場合の語形変化の様子を示す。この場合、与位格語尾の全候補二種類がそのまま接続した形の句候補二種類の他に、与位格固有のつなぎの子音字の挿入、および、つなぎの母音字の挿入が適用され、さらに二種類の句候補が生成される。そして、コーパスに対してこれらの句候補の検証を行なった結果、一種類の句候補のみが得られる。

表15の動詞の活用「副動詞:並列」の例、および、表16の動詞の活用「命令・願望:1-3 人称懸念」の例においては、「*арилга*」(消す)という動詞の活用の様子を示す。表15

の場合は、副動詞:並列の全語尾候補二種類がそのまま接続した形の句候補二種類が生成され、コーパスに対してこれらの句候補の検証を行なった結果、一種類の句候補のみが得られる。一方、表 16 の場合は、命令・願望:1-3 人称懸念の全語尾候補二種類がそのまま接続した後、母音字消失規則が適用された形の句候補二種類が生成され、コーパスに対してこれらの句候補の検証を行なった結果、一種類の句候補のみが得られる。

6. 実験および評価

名詞・動詞それぞれ 25 単語について、前節で述べた手順に従って、句候補を生成する実験を行なった。表 17、および、表 18 に、名詞・動詞それぞれの場合について、語尾あるいは活用の分類ごとに、語尾種類数、および、生成された句候補数の平均（一名詞もしくは一動詞あたり）を求めた結果を示す。さらに、コーパスに対する検索を行なう前の段階までで生成された句候補の中に正しい句候補を含む率、および、その正しい句候補がコーパスから検索できる率をそれぞれ示す。

評価結果を名詞と動詞の間で比較すると、特に、正しい句候補を含む率において、動詞の性能がかなり高いことが分かる。しかも、句候補数の平均も 1.22 と小さいため、コーパスを用いた絞り込みを行なわなくても、一意に近い活用変化が実現できていることが分かる。ただし、正しい句候補がコーパスに含まれる率は約半数と低い。特に、動詞の活用のうち、新聞でもよく用いられる一般的な活用はよく網羅されているが、特殊な活用が含まれる率が極端に低い。したがって、現時点で利用可能なコーパスを用いて、動詞の活用変化を網羅することは容易ではないと言える。動詞の場合は、活用変化の規則をより精密にすることにより、句候補生成の性能を高めるという戦略が現実的であると考えられる。一方、名詞については、語形変化における規則性がまだ十分に分析できていないため、句候補の網羅性が低くなっている。ただし、正しい句候補がコーパスに含まれる率は動詞よりは高く、句候補の絞り込みにおいてコーパスを利用する効果が、ある程度は期待できると言える。

7. 関連研究

文献 [2] においては、スペイン語およびルーマニア語について、時制・人称・数の活用変化形の生成規則を手で記述しておき、コーパス中で実際に観測される不規則変化形との間の類似度を計算して、各々の不規則変化形に対して、最も近い規則変化形の時制・人称・数を割り当てるという方法により、各言語での不規則変化形の形態素解析規則を獲得している。少数の語形変化規則およびコーパスを用いて、名詞・動詞の大規模な語形変化規則を獲得するという考え方は、本研究の手法の考え方に近いと言える。

8. おわりに

本稿では、現時点で利用可能なモンゴル語の言語資源、

特に、名詞・動詞の語幹のリスト、および、名詞・動詞に接続する語尾のリストから、モンゴル語の名詞句・動詞句を生成する手法を提案した。具体的には、名詞・動詞の語幹に語尾が接続する際の音韻論的・形態論的制約を整備し、語幹・語尾の語形変化の規則を作成した。評価実験の結果においては、8 割以上の場合について、生成された名詞句・動詞句の中に正しい句候補が含まれるという性能が達成できた。

文 献

- [1] С.Ганболор, Л.Тунгалаг, Зөв бичих дүрмийн тулгуур дохио, Улаанбаатар, 2000.
- [2] S. Cucerzan and D. Yarowsky. Bootstrapping a multilingual part-of-speech tagger in one person-day. In *Proc. 6th Conf. on Natural Language Learning*, pp. 132–138, 2002.
- [3] 栗林均. モンゴル語. 亀井孝, 河野六郎, 千野栄一(編), 言語学大辞典, 第 4 巻, 世界言語編 (下-2), pp. 501–517. 三省堂, 1992.

表 17 名詞の句候補手順の評価

語尾の分類	語尾種類数	(一名詞あたり)		
		句候補数の平均	正しい句候補を含む率 (%)	正しい句候補がコーパスに含まれる率 (%)
属格	7	1.92	72.0	64.0
対格	2	4.00	84.0	80.0
与位格	3	1.32	68.0	60.0
奪格	4	1.04	76.0	52.0
造格	4	1.04	88.0	64.0
共同格	3	1.04	96.0	84.0
再帰所属	4	1.04	88.0	80.0
合計/平均	27	1.63	81.7	69.1

表 18 動詞の句候補生成手順の評価

		活用の分類	語尾種類数	(一動詞あたり)		
				句候補数の平均	正しい句候補を含む率 (%)	正しい句候補がコーパスに含まれる率 (%)
1	命令・願望	1 人称意思 1	3	1.12	100.0	60.0
2		1 人称意思 2	2	1.12	100.0	8.0
3		2 人称命令	0	1.00	100.0	72.0
4		2 人称勧告	4	1.12	92.0	44.0
5		2 人称催促	4	1.12	92.0	24.0
6		2 人称懇願	2	1.12	100.0	4.0
7		1-3 人称願望	4	1.12	92.0	36.0
8		1-3 人称懸念	2	1.12	92.0	4.0
9	叙述	現在・未来	4	1.12	60.0	56.0
10		単純過去	4	1.12	100.0	100.0
11		体験過去	4	1.12	100.0	88.0
12		伝聞過去	2	2.00	100.0	68.0
13	形動詞	完了	4	1.12	100.0	100.0
14		継続	4	1.12	92.0	36.0
15		予定	1	1.00	100.0	92.0
16		習慣	4	1.12	100.0	96.0
17		可能性	4	1.12	100.0	52.0
18	副動詞	連合	1	1.00	100.0	92.0
19		並列	2	2.08	100.0	100.0
20		分離	4	1.12	92.0	80.0
21		条件	8	2.24	100.0	56.0
22		継続	4	1.12	100.0	56.0
23		限界	4	1.12	100.0	36.0
24		即刻	4	1.12	100.0	24.0
25		随伴	4	1.12	100.0	12.0
26		付帯 1	2	1.12	100.0	12.0
27		付帯 2	4	1.12	100.0	4.0
合計/平均			89	1.22	96.7	52.3