

音韻論的・形態論的制約を用いたモンゴル語形態素解析

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

E-mail: {enkh, utsuro, sato}@pine.kuee.kyoto-u.ac.jp

あらまし 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を用いて、モンゴル語の名詞句・動詞句の形態素解析を行なう手法を提案する。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。そして、この規則を用いて、語幹・語尾の組とそこから生成される句を対応させる語形変化テーブルを作成し、このテーブルを参照することにより、名詞句・動詞句の形態素解析を行なう。評価実験の結果においては、語形変化テーブルに登録されている句については、形態素解析の結果得られる語幹・語尾の組合せの候補の中に、正しい解析結果が必ず含まれることが確認できた。

キーワード モンゴル語, 形態素解析, 音韻論, 形態論, 語形変化

Mongolian Morphological Analysis based on Phonological and Morphological Constraints

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

[†] Dpt. Intelligence Sci. and Tech., Graduate School of Informatics, Kyoto University

E-mail: {enkh, utsuro, sato}@pine.kuee.kyoto-u.ac.jp

Abstract 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 for morphologically analyzing noun/verb phrases of the Mongolian language. More specifically, we first examine phonological and morphological constraints on connecting stems of nouns/verbs and suffixes, and invent inflection/conjugation rules for nouns/verbs. Then, we compile a table for mapping a stem-suffix pair and a phrase to be generated from the stem-suffix pair. Morphological analysis of noun/verb phrases is performed by simply consulting this mapping table. We experimentally show that, by the proposed method, correct candidate of stem-suffix pair can be obtained from the given noun/verb phrases.

Key words Mongolian language, morphological analysis, phonology, morphology, inflection/conjugation

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。また、そのような句生成のための言語知識・規則を運用すれば、モンゴル語の句の形態素解析を行なうこともできるが、現時点では、モンゴル語文の形態素解析を実用的規模で行なうことも実現されていない。

このような状況をふまえて、我々は、[4]において、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。ここでは、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。また、評価実験の結果において、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を達成した。[4]の結果をふまえて、本稿では、モンゴル語の名詞句・動詞句の形態素解析を行なう手法を提案する。具体的には、まず、既存のモンゴル語辞書から名詞語幹および動詞語幹を人手で抽出する。本稿の実験においては、名詞語幹 1,926 語、および、動詞語幹 1,254 語を抽出した。次に、これらの語幹に対して、[4]において整備したモンゴル語名詞句・動詞句生成規則を適用することにより、語幹・語尾の組から句を生成するための語形変化テーブルを作成する。そして、この語形変化テーブルを参照することにより、与えられた名詞句・動詞句を形態素解析して語幹・語尾に分離する。評価実験の結果においては、語形変化テーブルに登録されている句については、形態素解析の結果得られる語幹・語尾の組合せの候補の中に、正しい解析結果が必ず含まれることが確認できた。

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、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. で述べる語幹・語尾の接続制約を満たす語尾が選ばれる。さらに、4. 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^(注1)。

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. 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. 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^(注2)。

なお、名詞・動詞に関して、本稿の執筆段階において実装されていないものとして、派生語がある。派生語とは、名詞語幹

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
対格	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
xyqaxd (子供)
- (2) 名詞語幹 + 複数語尾
xyqaxdyqd (子供達)
- (3) 名詞語幹 + 複数語尾 + 格語尾
xyqaxdyqdməy (子供達と一緒に)
- (4) 名詞語幹 + 複数語尾 + 格語尾 + 再帰所属語尾
xyqaxdyqdməygeə (自分の子供達と一緒に)
- (5) 名詞語幹 + 複数語尾 + 否定の語尾 + 格語尾 + 再帰所属語尾
xyqaxdyqdeygeəpəə (自分の子供達とは別に)

図 1 名詞の語形変化の例

あるいは動詞語幹に派生語語尾が接続して語形変化した結果の語であり、名詞語幹から構成される派生動詞、名詞語幹から構成される派生名詞、および、動詞語幹から構成される派生名詞がある。派生語語尾としては、数十種類のものがある。派生語の内部構造を解析するためには、派生語に対して語幹・語尾の語形変化規則を適用して形態素解析を行なう必要があるが、実用的には、既知の派生語を語幹として登録し、形態素解析を行なうという方式が妥当であると考えられる。名詞・動詞の派生語については、今後実装する予定である。また、句生成および形態素解析において、名詞と同様の扱いが可能なものとして形容詞がある。モンゴル語の形容詞には、名詞に接続する語尾のうち複数語尾を除く語尾がすべて接続可能であり、語幹・語尾の語形変化の規則についても、名詞句の語形変化で用いている規則がそのまま適用できる。形容詞句の生成および形態素解析についても今後実装する予定である。

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。この語尾の接続においては、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、一定の接続制約が満たされる必要がある。これらの制約の詳細については、[4]を参照されたい(なお、[4]で述べた内容は文献[1]に基づいており、日本語での用語等は文献[5]に従っている。)

[4]で述べた制約と比べると、本稿における制約においては、中性母音 *u* の優先度を下げ、男性母音・女性母音を優先するという部分が改善されている。これにより、句生成の候補数を減らすことができおり、6. で述べるように、句候補生成の性能が向上した。また、本稿では、動詞・形動詞・予定形に対して

(注1): このうち、本稿の執筆段階では、複数語尾、および、否定の語尾が未実装である。これらについては、今後実装する予定である。

(注2): このうち、本稿の執筆段階では、使役態語尾(数種類)、受動態語尾(一種類)、否定の語尾、および、完了の語尾(形動詞・完了形よりも完了の意味が強い)が未実装である。これらについては、今後実装する予定である。

表 3 動詞語幹+形動詞・予定形語尾+与位格語尾+再帰所属語尾の語形変化の例

動詞語幹	<i>авар</i> (救う)
語形変化	<i>авар</i> + <i>x</i> (形動詞・予定形語尾) → <i>аврах</i> (救うこと) <i>аврах</i> + <i>ыг</i> (与位格語尾) → <i>аврахыг</i> (救うことを) <i>аврахыг</i> + <i>аа</i> (再帰所属語尾) → <i>аврахыгаа</i> (自分が救うことを)

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

適用する母音字の接続制約について、動詞・形動詞・予定形に特化したものを用いている。これにより、句生成の精度が[4]よりも改善した^(注3)。なお、本稿の執筆段階では未実装であるが、母音字の接続制約において、一つ前に接続した語尾の分類ごとに接続制約を細分化することにより、制約を厳しくすることができ、誤った句候補の生成が抑えられることが期待できることが分かっている。今後は、この考えに基づいて、母音字の接続制約を細分化する予定である。

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

(注3)：動詞・形動詞・予定形は、文法的には名詞としての変化をする。表3に、動詞語幹+形動詞・予定形語尾+与位格語尾+再帰所属語尾の語形変化の例を挙げる。動詞・形動詞・予定形に接続する語尾は、表1の一覧よりも少なく、全体で20種類である。

- (1) 動詞語幹
уд (食べる)
- (2) 動詞語幹 + 受動態語尾
удэгд (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
удье (食べよう)
- (5) 動詞語幹 + 単純過去の語尾
удсэн (食べた)
- (6) 動詞語幹 + 形動詞・完了の語尾 + 否定の語尾
удсэнгүй (食べなかった)
- (7) 動詞語幹 + 従属節 (限界) 語尾
удтэл (食べるまで)
- (8) 動詞語幹 + 受動態語尾 + 単純過去の語尾
удэгдсэн (食べられた)

図 2 動詞の活用の例

の複数の可能性の中から、まず、前節で述べた母音字の接続制約を満たす語尾が選ばれ、さらに、語幹・語尾の接続制約を満たす語尾が選ばれる。そして、以下の四種類の語形変化規則により、語幹・語尾が語形変化する。

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

これらの制約・規則の詳細については[4]を参照されたい。

なお、本稿で実際に用いた制約・規則を、[4]で述べた制約・規則と比べると、特に、動詞・形動詞・予定形に対して適用する制約・規則の部分が異なっている。具体的には、動詞・形動詞・予定形に対しては、つなぎの母音字の挿入規則の一部のみを用いる。これにより、句生成の精度が[4]よりも改善した。また、本稿の執筆段階では未実装であるが、3.で述べた母音字の接続制約の場合と同様に、本節の語幹・語尾接続制約・語形変化規則の全体においても、一つ前に接続した語尾の分類ごとに、制約・規則を細分化することにより、誤った句候補の生成が抑えられることが期待できることが分かっている。今後は、この考えに基づいて、制約・規則を細分化する予定である。

5. 名詞・動詞の語幹リストの作成

見出し語数約 7,500 語の日本語・モンゴル語対訳辞書^(注4)のモンゴル語見出し語から、以下の手順で、名詞・動詞の語幹を抽出した。まず、名詞については、見出し語が名詞の語幹で記述されているので、1,926 語を手で抽出した。一方、動詞に

(注4)：制作者の清水幹夫氏から提供して頂いた。

表 4 名詞+語尾の語形変化テーブルの例 1

語幹/語幹品詞	боловсрол(教育)/名詞
語尾/語尾種類	ын(~/の)/属格
語形変化後の語	боловсролын(教育の)

表 5 名詞+語尾の語形変化テーブルの例 2

語幹/語幹品詞	боловсрол(教育)/名詞
語尾/語尾種類	д(~/に)/与位格
語形変化後の語	боловсролд(教育に)

については、見出し語が形動詞・予定形で記述されている。そこで、まず、動詞の形動詞・予定形 1,254 語を手で抽出し、形動詞・予定形を動詞・語幹と予定形・活用語尾に分離する形態素解析規則を適用した。この形態素解析における語幹の候補語数は、形動詞・予定形一単語あたり、平均で 1.365 語であり、この中に正しい語幹を含む率は 100%であった。この形態素解析結果に対して、手で正しい語幹を選択し、動詞の語幹リストを作成した。

6. モンゴル語句候補生成の評価

前節で作成した、名詞語幹 1,926 語、および、動詞語幹 1,254 語について、以下の手順で名詞句・動詞句の句候補生成を行ない、その性能を評価した。

(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。

ただし、名詞については、表 1 中の再帰所属を除く 23 種類の格語尾のみを評価対象とし、名詞の語幹にこれらの語尾が一つだけ接続した句を生成する過程を評価した。動詞についても、動詞の語幹に表 2 の活用語尾が一つだけ接続した句を生成する過程を評価した。評価実験の結果では、名詞については、句候補の平均数が 1.68、正しい句を含む率が 98.86%であった。動詞については、句候補の平均数が 1.01、正しい句を含む率が 99.98%であった。この結果から分かるように、正しい句を含む率については、ほぼ 100%に近い性能を達成することができた。

また、動詞・形動詞・予定形 1,254 語を名詞とみなして、名詞句候補生成の評価を行なった。ここでは、動詞・形動詞・予

表 6 動詞+活用語尾の語形変化テーブルの例 1

語幹/語幹品詞	арилга(消す)/動詞
語尾/語尾種類	ж(~/して)/副動詞:並列
語形変化後の語	арилгаж(消して)

表 7 動詞+活用語尾の語形変化テーブルの例 2

語幹/語幹品詞	арилга(消す)/動詞
語尾/語尾種類	уузай(~/するな)/命令・願望:1-3 人称懸念
語形変化後の語	арилгуузай(消すな)

定形に対して、表 1 中の再帰所属を除く 23 種類の格語尾が接続した句、および、それらの句に対して、再帰所属語尾が接続した句を生成する過程を評価した。ただし、3. および 4. で述べたように、母音字の接続制約、語幹・語尾の接続制約、および、語幹・語尾の語形変化規則については、動詞・形動詞・予定形に特化したものを用いた。この結果においては、句候補の平均数が 1.01、正しい句を含む率が 100%となり、動詞・形動詞・予定形に特化した制約および規則が有効であることが検証できた^(注5)。

7. 語幹・語尾の語形変化テーブルの作成

5. で作成した名詞語幹 1,926 語、および、動詞語幹 1,254 語について、以下の手順で語幹・語尾の語形変化テーブルを作成した。まず、前節の手順により、句候補を生成した後、正しい句を手で選択した。ただし、名詞句としては、名詞の語幹に対して、表 1 中の再帰所属を除く 23 種類の格語尾が一つだけ接続した句を対象とした。また、動詞句としては、動詞の語幹に表 2 の活用語尾が一つだけ接続した句を対象とした。さらに、動詞・形動詞・予定形を名詞とみなして生成した句も対象とした。ここでは、動詞・形動詞・予定形に対して、表 1 中の再帰所属を除く 23 種類の格語尾が接続した句、および、それらの句に対して、再帰所属語尾が接続した句を対象とした。

次に、語幹・語尾、および、語形変化後の句の情報を用いて語形変化テーブルを作成した。語形変化テーブルは以下の情報から構成した。

- 語幹、もしくは、語幹にいくつかの語尾が接続して語形変化した語、および、語幹の品詞。
- 新たに接続する語尾の種類、および、語尾。
- 語形変化後の語。

語形変化テーブルの例を表 4~7 に示す。名詞語幹 1,926 語、および、動詞語幹 1,254 語に対して、語形変化テーブルに登録された名詞句の数は、それぞれ、10,529 個、および、29,632 個となった。また、動詞・形動詞・予定形 1,254 語に対して、語形変化テーブルに登録された句の数は、16,302 個となった。

なお、語形変化テーブルを用いて句の形態素解析を行なった結果、語幹・語尾の組合せとして複数の候補が得られる場合がある。これらの例を表 8 に示す。一つ目の例においては、“op(ベッド、代わり)”あるいは“орон(国)”という、異なる二つの名詞語幹に対して、“ны(~/の)”あるいは“ы(~/の)”という異なる属格語尾が接続して語形変化した結果の句が同じ表記になっている。この例の場合は、文の意味を考慮して形態素解析の曖昧性を解消する必要がある。一方、二つ目の例においては、“хаз(囁む)”という動詞語幹に対して、叙述・単純過去形語尾

(注5)：[4] においては、名詞・動詞それぞれ 25 語について、句候補生成の評価を行なった。名詞については、句候補の平均数が 1.63、正しい句を含む率が 81.7%であった。動詞については、句候補の平均数が 1.22、正しい句を含む率が 96.7%であった。この評価実験と比べると、本稿で行なった評価実験においては、3. および 4. で述べた改善が追加されており、さらに、プログラム上の誤りが除去されている。これらにより、本稿で行なった評価実験においては、名詞の句候補に正しい句を含む率が大幅に向上し、動詞についても、句候補に正しい句を含む率が向上した。

表 8 形態素解析において複数の解析結果が得られる例

句	形態素解析結果 (語幹/語幹品詞+語尾/語尾種類)
<i>орны</i>	<i>ор</i> (ベッド, 代わり)/名詞 + <i>ны</i> (~の)/属格 <i>орон</i> (国)/名詞 + <i>ы</i> (~の)/属格
<i>хазсан</i>	<i>хаз</i> (嘸む)/動詞 + <i>сан</i> (~した)/叙述・単純過去 (嘸んだ(文末)) <i>хаз</i> (嘸む)/動詞 + <i>сан</i> (~した(連体修飾))/形動詞・完了 (嘸んだ(犬))

表 9 コーパス中の句の内訳 (% (個数))

語形変化テーブル中の 語幹・句の有無	名詞	動詞	形動詞 ・予定形	その他	合計
語幹・句とも存在する	42.1 (74)	36.5 (62)	56.1 (32)	0 (0)	24.7 (168)
語幹のみ存在する	25.0 (44)	29.4 (50)	29.8 (17)	0 (0)	16.3 (111)
語幹・句とも存在しない	32.9 (58)	34.1 (58)	14.1 (8)	100 (277)	59.0 (401)
合計	25.9 (176)	25.0 (170)	8.4 (57)	40.7 (277)	100 (680)

あるいは形動詞・完了形語尾が接続しているが、この二つの語尾が同じ表記となっており、語形変化した結果の句も同じ表記となっている。叙述・単純過去形の場合は、文末等に現れる過去形となり、形動詞・完了形の場合は、連体修飾用法となる。この例の場合は、この句の直後が名詞句かどうかによって、形態素解析の曖昧性を解消することができる。

8. モンゴル語形態素解析の評価実験

モンゴル語形態素解析の評価実験を行なうために、まず、モンゴル語コーパスを収集した。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。このコーパスから、無作為に 680 語を収集し、前節で用意した動詞・名詞の語形変化テーブルを用いて各語の形態素解析を行なった。680 語のうち、語形変化後の語が語形変化テーブルに含まれる語については、形態素解析の結果得られる語幹・語尾の組合せの候補の中に、正しい解析結果が必ず含まれていた。また、この 680 語において、複数の形態素解析結果が得られたのは 9 語のみであり、そのいずれも、表 8 の二つ目の例と同様の、動詞・叙述・単純過去形と動詞・形動詞・完了形の間の曖昧性であった。今回の評価実験では、名詞の複数語尾が未実装であったため、複数の形態素解析結果が得られる割合は少なかったが、今後、名詞の複数語尾を実装すると、複数の形態素解析結果が得られる割合が増えるものと思われる。

次に、前節で用意した動詞・名詞の語形変化テーブルが、コーパス中のどの程度の範囲の語に対応しているかのカバレッジを評価するために、まず、680 語を、名詞、動詞(形動詞・予定形を除く)、動詞の形動詞・予定形、その他の単語に分類し、それぞれのクラスについて、語形変化テーブルに含まれるかどうかを判別し、以下に分類した。

- 「語幹・句とも存在する」
- 「語幹のみ存在する」
- 「語幹・句とも存在しない」

「語幹・句とも存在する」は、コーパス中の出現形が、そのまま

の形で、語形変化後の語として語形変化テーブルに含まれるものである。「語幹のみ存在する」は、コーパス中の出現形から判別した語幹は語形変化テーブルに含まれるが、コーパス中の出現形が、そのままの形で、語形変化後の語として語形変化テーブルに含まれてはいない、というものである。これらの語については、2. で述べたように、今後、名詞の複数語尾・否定の語尾、および、動詞の使役態語尾・受動態語尾・否定の語尾・完了の語尾が接続する場合の語形変化を実装することにより、形態素解析が可能となる。「語幹・句とも存在しない」は、コーパス中の出現形から判別した語幹が語形変化テーブルに含まれない、というものである。これらの語については、語形変化テーブルに登録する語幹そのものを増やさなければ、形態素解析することができない(これらの語の中には、語形変化テーブルに含まれる語幹から派生した語も含まれる。)

9. 関連研究

[3] においては、日本語形態素解析システム茶筌^(注6)を処理系として、モンゴル語文の形態素解析を行なうための文法体系の構築を試みている^(注7)。茶筌の処理系は、基本的には、活用語・非活用語の品詞体系、および、活用語の活用語尾の語形変化の体系を定義する機能を持つ。また、各形態素、および、二個もしくは三個程度の形態素の接続に対してコストを定義することにより、形態素解析の結果得られる複数の解析結果を絞り込む機能を持つ。[3] においては、茶筌の処理系が持つ機能のうち、活用語の活用語尾の語形変化の体系を定義する機能を利用することにより、名詞・動詞の詳細な活用型・活用形を定義している。また、語尾については、活用語とはせず、変化形をすべて別形態素として登録している。この方式と比較すると、本稿のアプローチは、茶筌辞書のような明示的な文法体系を立てるのではなく、できる限り抽象化したレベルで音韻論的・形態論的特性を整備し、この制約を用いて語幹・語尾の接続制約・語形変化規則を記述するというアプローチであると言える。本

(注6) : <http://chasen.naist.jp/hiki/ChaSen/>

(注7) : 本節の議論にあたって、[3] の辞書データを提供して頂いた諏訪東京理科大学 江原暉将先生に感謝する。

稿のアプローチでは、形態素として辞書に登録されるのは、(名詞・動詞の) 語幹およびそれらの語幹に接続する語尾(の基本形)のみとなり、語幹や語尾の変化形を別途登録することはない。そのかわりに、語幹・語尾が語形変化して生成される句については、その全ての可能性を語形変化テーブルに登録することとなる。ここで、日本語文の形態素解析と、モンゴル語における句の形態素解析を比較すると、モンゴル語においては、句が空白により分かち書きされる点が特徴的である。したがって、モンゴル語の句の形態素解析においても、実用的には、語幹に対して高々数個の語尾が連続して接続する可能性を考慮すれば十分である。本稿では、モンゴル語におけるこの特性を考慮して、句の候補をすべてテーブルに登録するアプローチを採用している。

[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。

10. 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を用いて、モンゴル語の名詞句・動詞句の形態素解析を行なう手法を提案した。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。そして、この規則を用いて、語幹・語尾の組とそこから生成される句を対応させる語形変化テーブルを作成し、このテーブルを参照することにより、名詞句・動詞句の形態素解析を行なう手法を提案した。評価実験の結果においては、語形変化テーブルに登録されている句については、形態素解析の結果得られる語幹・語尾の組合せの候補の中に、正しい解析結果が必ず含まれることが確認できた。

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] 江原暉将, 早田清冷, 木村展幸. 茶釜を用いたモンゴル語の形態素解析. 言語処理学会第10回年次大会論文集, pp. 709-712. 言語処理学会, 2004.
- [4] Sanduijav ENKHBAYAR, 宇津呂武仁, 佐藤理史. 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, Vol. 2004, No. (2004-NL-162), pp. 87-93, 2004.
- [5] 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.