

モンゴル語-日本語翻訳のための支援ツールに関する検討

大平 栄二* 上谷 恵里奈** 河本 美月**

(2019年1月7日受理)

1. はじめに

モンゴル語はいわゆる「アルタイ語族」に属し、文法構造の面で日本語に類似するところが多い。例えば、日本語と同様に語順類型が SOV 型であり、必ず修飾語が被修飾語の前に置かれる。また、語幹に語尾をつけてどんどん拡大していくタイプの言語であり、類型的にも日本語と同じ膠着言語である。膠着言語とは、文法的な意味を表す助詞や活用語尾などの接辞が、実質的な意味を表す名詞や活用語の語幹などの語に膠着する言語のことである。

このように、両者には共通点が多く、モンゴル語は日本人にとって学びやすい言語であると思われるが、他の言語と同様に、複雑な正書法があり、習得が容易ではないことが知られている[1]。例えば、モンゴル語は、英語などと同様に単語（正確には日本語の文節に相当）がスペースなどで分かち書きされるため、日本語に比べ処理が容易である。本論文では、以降この分かち書きされた単語をワード[2]と呼ぶことにする。しかし、モンゴル語の母音には、短母音、長母音、さらに、男性や女性、中性母音など様々な種類があり、名詞と格語尾が接続してワードを作る際、母音調和（ワード内に現れる母音の組み合わせの制約）により、接続する語尾が変化する。さらに、母音の挿入や削除[1,3]も加わり、特に初心者にとっては、単語辞書で単語を検索できない場合も少なくない。これが、モンゴル語学習の大きな壁の一つになっている。

モンゴル語などの言語リソースが十分でない言語の研究や学習においては、言語ツールが有効であるとの報告[4]がある。現在、インターネット上で利用できる単語辞書は増えつつある。しかし、語幹や語尾の変形に対応した、モンゴル語-日本語の検索ツールは現時点で存在しない。モンゴル語-日本語の自動翻訳を目指した研究[5-7]が進められているが、他の言語と同様に、すぐには実現困難であると考えられる。また、この技術をイン

ターネット上で利用可能とするには多くの工数や費用を要する。

われわれは、このような問題を解決して、インターネット上などで、人が手軽にモンゴル語文章を日本語に翻訳するのを支援するツールの開発を進めている[8,9]。本論文では、モンゴル語から日本語への翻訳に特化することにより、CGIを実現するプログラム言語 Perlなどで実装できる簡易な変換規則とそれに伴う単語辞書を提案し、その有効性を検証する。以下、2章でモンゴル語の名詞の特徴、3章で動詞の特徴について説明する。そして4章で、その特徴に基づいたモンゴル語-日本語変換法を提案し、情報関係の専門テキストを用いて行った検証結果について述べる。

2. モンゴル語名詞の特徴

2.1 モンゴル語文法

モンゴル語には、様々な文字や方言があるが[10,11]、本論文では、現在、モンゴル国で用いられているキリル文字表記のハルハ方言を対象としている。モンゴル語の表記は、ロシア語のアルファベットにモンゴル語特有の2文字「ө」と「ү」を加えた35文字であり、母音12字、子音20字、半母音1字（й）の計33文字と記号文字「ь（硬音符）」と「ъ（軟音符）」の2つからなる[3,11]。母音と子音を以下に示す。

母音：а, о, у, э, ө, ү, и,

е, ю, я, ё, ы

子音：б, в, г, д, ж, п, р, с, т, ф,

з, к, л, м, н, х, ц, ч, ш, щ

このうち、а, о, у, э, ө, ү, иの7つが短母音で、短母音を2つ並べると長母音となる。たとえば、аа, ооである。ただし、иの長母音は、ийまたはыの1文字で書かれる。また、二重母音は、ай, ой, уй, үйの4つで、短母音において、а, о, уが男性母音、ү, э, өが女性母音、иが中性母音である。

モンゴル語と日本語は、前述したように、語順類型が SOV 型である点が共通している。他の言語、例えば英語（語順類型は SVO 型）との比較のため、以下に例を示す。

* 電子情報工学科

** 元津山工業高等専門学校 情報工学科

(例)

モンゴル語： би тэр номыг авав

日本語： 私は あの 本を 買った

英語： I bought that book

上記において би は「私」、тэр は「あの」、номыг は「本を」、авав は「買った」に対応する。述語が必ず文の最後に置かれ、修飾語は必ず被修飾語の前に置かれる。また、モンゴル語と日本語は膠着言語であり、名詞などの語幹に、格を表す語尾（「は」や「を」）や数を表す語尾が接続され名詞句を形作る。以下に、その例を示す。

「номыг」本を

= 語幹「ном」本 + 「ыг」を

「өдрийн」日の

= 語幹「өдөр」日 + 「ийн」の

2. 2 名詞とその変形

(1) 名詞の語尾

前述したように、モンゴル語は、語幹にさまざまな語尾を付けて名詞などのワードを形作る膠着言語である。表 1 に代表的な表層格の観点から名詞の変形について整理した結果を示す。モンゴル語は、日本語と異なり、語尾が接続すると名詞や動詞の語幹の形が変形することがある。また、逆に語幹により語尾が変形することがある[3]。その変形は、一定の規則にしたがって起こる。

表 1 において、○印は、項目に対応する語幹の変形が生じることを示す。また、点線の上段は原型の語尾、下段が変形後の語尾を示す。一部例を説明すると、日本語では主格助詞として、「は」または「が」が使われるが、モンゴル語では、主格の語尾は、形のないゼロ語尾である。すなわち、語幹には何も付かない。この点は英語と似ている。対格もゼロ語尾になる場合がある。

同じ奪格でも、名詞の母音構造に応じて使う語尾を「аас, оос, ээс, өөс」の中から選択して変える必要がある。以下に例を示す。これが母音調和と呼ばれるもので、翻訳を複雑にしている。

төхөөрөмжөөс (設備から)

программаас (プログラムから)

ここで、モンゴル語では、主語との関係を表す再帰語尾「аа, оо, ээ, өө」があり、表 1 の語尾の後に追加されるが、対格語尾の後に再帰語尾が置かれる場合、対格語尾が省略される場合がある[3]。このため、文法的には適切ではないが、再帰語尾も対格の語尾として扱うことにした。

なお、方位格は他の語尾と異なり、名詞に接続するのではなく、後置詞としてスペースで区切られた 1 つのワードとして表記される。英語の前置詞と似た表記である。

表-1 名詞の語尾の一覧

Table 1 List of particles of the noun

| 分類 | モンゴル語 | 日本語 | 交代 | | 脱落 | 追加 |
|-----|---|-----------------|----|---|----|----|
| | | | Г | Н | | |
| 主格 | なし | は/が | | | | |
| 対格 | ыг, г, ийг, (аа, ээ, оо, өө) | を | ○ | | ○ | |
| | (гаа, гээ, гоо, гөө, а, э, о, ө) | | | | | |
| 属格 | ын, ы, н, ийн, йн | の | ○ | ○ | ○ | |
| 与位格 | д, т, нд | に | | ○ | | ○ |
| | ад, аг, эд, эт, од, от, өд, өт нд, нг анд, энд, онд, өнд, инд | 時に | | | | |
| 奪格 | аас, оос, ээс, өөс | から | ○ | ○ | ○ | |
| | гаас, гоос, гээс, гөөс ас, ос, эс, өс | | | | | |
| 造格 | аар, оор, ээр, өөр | で | ○ | | ○ | |
| | гаар, гоор, гээр, гөөр ар, ор, эр, өр | | | | | |
| 共同格 | тай, той, тэй (正書法で той は無い) | と/ もつ/ ある | | | | |
| 欠如格 | гүй | 無で | | | | |
| 方向格 | руу, луу, рүү, лүү, уруу, үрүү | へ | | | | |

(2) 語尾の変化

先頭が長母音からなる奪格や造格および再帰語尾のような語尾は、語幹によって表 1 の点線の下語尾に変化することがある。例えば、語尾「аас」を例にとると、先頭が短母音になった「ас」(語幹末尾が「и」や「я」の場合)や先頭に文字「Г」が接続された「гаас」(語幹末尾が「и」以外の短母音の場合)も用いられる[3]。このため、これらも語尾として登録することにした。

(3) 語幹の交代

モンゴル語では、特定の語尾を接続するときだけ現れる別の形の語幹を交代語幹[3]と呼ぶ。交代語幹には、「хөдөөн」や「уулан」のように末尾に「н」が出てくる Н交代語幹と、「санг」のように末尾に

「r」が出てくる r 交代語幹とがある。

(a) H 交代の例

「хөдөөнөөс」 田舎から
 = 「хөдөө」 田舎 (H 交代語幹 : хөдөөн)
 + 「өөс」 から

ただし、以下に示すように、H 交代語幹を持つにも関わらず、属格の場合にだけ交代しない語がある。このように、不規則な活用をするため、ひとつずつ覚えるほかないようである。

「номын」 本の
 = 「ном」 本 (H 交代語幹 : номон)
 + 「ын」 の

(b) r 交代の例

「сангаас」 募金から
 = 「сан」 基金 (r 交代語幹 : санг)
 + 「аас」 から

対格の語尾「r」が接続するときに、r 交代語幹の末尾の「r」は脱落する。すなわち、元の語幹にそのまま語尾「r」を接続したのと同じ綴りになる。例を示す。

「дэнг」 ランプを
 = 「дэн」 ランプ (r 交代語幹 : дэнг)
 + 「г」 を (r 脱落)

交代語幹[3,12]は、古いモンゴル語の形の名残である。現代のモンゴル語では、ふつうの語幹ではなく交代語幹を使うことによって、語の意味を区別したりする新たな用法が発達しているようである[3]。

(4) 末尾母音の脱落

対格、属格、奪格、造格語尾が付く際には、末尾の母音が脱落する[3]。これは、末尾母音の脱落と呼ばれ、「и」以外の短母音で終わる単語に長母音で始まる語尾が接続される際は、最後の母音が脱落する。用法の例を以下に示す。

「хөрөнгийн」 財産の
 = 「хөрөнгө」 財産
 + 「ийн」 の (ө 脱落)

(5) 未聞母音の脱落

未聞脱落とは、子音で終わる単語に、長母音で始まる語尾が接続される際に、末尾(子音)の前の母音である未聞母音が脱落する変形である[12]。用法の例を以下に示す。

「олноос」 多数より
 = 「олон」 多数
 + 「оос」 より (о 脱落)

(6) 末尾短母音の追加

語尾の接続により、文字の脱落のみでなく、文字が追加される場合がある。与位格語尾が付く際に追加される。このため、語尾として、この追加される母音と語尾を合わせた表 1 に示す語尾も登録している。

「улсад」 国に
 = 「улс」 国
 + 「д」に (a 追加)

(7) 複数形

名詞語幹に「ууд」か「үүд」を接続することにより複数形となる。複数形をとる場合も、前述した未聞母音や末尾短母音の脱落が生じることがある。

「хэлүүд」 言語(複数)
 = 「хэл」 言語
 + 「үүд」

(8) 複合語

モンゴル語でも他の言語と同様、複数単語で元の単語と異なった意味を表す単語となる複合語がある。例えば、「санах(覚える)」と「ой(記憶)」の2語で、記憶装置「санах ой」という単語となる。したがって、翻訳支援には複合語の処理が不可欠である。複合語では、語尾は最後の単語の語幹に接続する。このため、複合語は、上記の例の様に、単語間に小文字のスペースを挟んだ形で1つの単語として辞書に登録する。また、今回作成した辞書には、2語の複合語が約6000単語、3語が約600単語、4語以上が50単語あったが、4語以上の複合語は、ほとんどが説明文の形態であった。このため、3語までの複合語を処理の対象とした。

3. モンゴル語動詞の特徴

3.1 動詞の構造

辞書には、動詞の語幹に「х, ах」などが接続されたものが見出し語として登録されている。これを不定形と呼ぶ。以下に例を示す。ハイフン(-)の前が語幹である。ここで、ハイフンは、分かりやすくするためにつけたもので、実際の単語には付かない。

гүй-х 走る алх-ах 歩く
 унш-их 読む ид-эх 食べる

動詞のワードも日本語と仕組みが似ており、「動詞語幹」+「補助語幹」+「語尾」で構成される[3]。日本語の動詞「する」が、「した」、「している」、「すれば」と活用するように、モンゴル語も語幹に語尾を接続することにより、さまざまな形に活用し変形する。「食べる」の不定形「идэх」を例に説明する。

| | | |
|----|---------|------|
| 現在 | иднэ | 食べる |
| 過去 | идсэн | 食べた |
| 否定 | идэхгүй | 食べない |
| 仮定 | идвэл | 食べれば |

モンゴル語では、接続する語尾により、終止形、希求形、副動詞形、形動詞形の4種類の形のうちの1つを必ずとる[3]。表2に終止形と希求形、表3に副動詞形の語尾を示す。表1と同様に、点線の上段は原形の語尾、下段が変形後の語尾を示す。2.2節

表-2 動詞の語尾の一覧

Table 2 List of particles of the verb

| 分類 | モンゴル語 | 日本語 |
|-----------|--|-------------|
| 過去 | сан, сэн, сон, сөн, жээ, чээ, в, лаа, лээ, лоо, лөө ----- ав, эв, ов, өв, ив | ～した |
| 現在・未 来 | на, нэ, но, нө ----- ана, энэ, оно, өнө, ина, инэ, ино, инө | ～する |
| 習慣 | даг, дэг, дог, дөг ----- | いつも～する |
| 否定 | хгүй, аагүй, ээгүй, оогүй, өөгүй, сангүй, сэнгүй, сонгүй, сөнгүй ----- гаагүй, гээгүй, гоогүй, гөөгүй, агүй, эгүй, огүй, өгүй, ахүй, эхүй, охүй, өхүй, ихүй | ～(し)ない |
| 意志 | я, ё, е, сугай, сүгэй, сугой, хаас, хээс, хоос, хөөс ----- ья, ья, ье, ье, ье, ье, ая, эя, оя, өя, | ～(し)よう |
| 願望・希 望 | аасай, ээсэй, оосой, өөсэй, маар, мээр, моор, мөөр тугай, тугэй, тугой ----- гаасай, гээсэй, гоосой, гөөсэй, асай, эсэй, осой, өсэй, | ～したい |
| 依頼・命 令 | аарай, ээрэй, оорой, өөрэй аач, ээч, ооч, өөч, ----- гаарай, гээрэй, гоорой, гөөрэй, арай, эрэй, орой, өрэй, гаач, гээч, гооч, гөөч, ач, эч, оч, өч, | ～してくださ い |

(2)で述べた名詞の語尾と同様に、動詞でも先頭が長母音からなる否定や願望・希望および依頼・命令の語尾は、語幹によって点線の下段に示す語尾に変化することがある。

さらに、動詞の場合、子音で始まる語尾が接続する時、動詞語幹の末間母音位置に短母音を追加される変形が生じることがある。

「хосолж」ペアになって
= 「хослох (хосл-)」ペアになる
+ 「ж」 (o 追加)

3. 2 終止形

述語として文を終わらすことができる形である。表2における過去、現在・未来、習慣および否定が該当し、語幹に接続する語尾によって時制やアスペクトを表す。

(1) 過去

モンゴル語では、過去を表す時制として一般過去、伝聞過去、体験過去、文書的な過去の4種類がある。

(a) 一般過去

「動詞語幹」+「сан (сэн, сон, сөн)」で構成される。用法例を以下に示す。なお、()の中は動詞の語幹を示している。

「ирсэн」来た
= 「ирэх (ир-)」来る
+ 「сэн」一般過去

(b) 伝聞過去

「語幹」+「жээ (чээ)」で構成され、人から聞いたことや歴史的な出来事などを表す。

「иржээ」来たようだ

(c) 体験過去

「語幹」+「лаа (лээ, лоо, лөө)」で構成され、話し手が自身で体験したときや見聞いたことを表す。

「сонслоо」聞いた
= 「сонсох (сонс-)」聞く
+ 「лоо」過去

「нүүх боллоо」引越しをするようになった

= 「нүүх (нүү-)」引越しする

+ 「боллоо (「бол」なる+「лоо」体験過去)

後者の例に示すように「不定形」+「боллоо」で、日本語の「何々をするようになった」という意味を表す。後述するが、このような場合には、不定形が使われる。

(d) 文書的な過去

「語幹」+「в」で構成され、比較的中立的で客観的な文章的過去または近い過去に起こった動作を表す。語幹が子音で終わるときは短母音が挿入される。

「харав」見た
= 「харах (хар-)」見る
+ 「в」過去 (a 追加)

また、語幹末がьで終わる場合は、語幹末のьの部分母音字иに変形する。

「урив」招いた
= 「урих (урь-)」招く
+ 「в」

(2) 現在・未来

モンゴル語では、現在と未来の区別はなく、過去ではないものは非過去と呼ばれ、語尾「на (нэ, но, нө)」が接続する。この語尾でも、語幹が子音で終わるときは短母音が挿入される場合がある。

Энэ загас мах иднэ. :
この魚は肉を食べる。(恒常的な動作)
「идэх (ид-)」食べる+「нэ」

Аав маргааш ирнэ. :
お父さんは明日来る。(未来)
「ирэх (ир-)」来る+「нэ」

(3) 習慣

習慣は、時間と関係ない一般的な出来事を表すときに用いられ、「語幹」+連体形の語尾「даг (дэг, дог, дөг)」で構成される。用法の例を以下に示す。

Би байнга номын санд хичээл хийдэг.

日本語：私はいつも図書館で勉強をする。

「хийдэг」(いつも) する
= 「хийх (хий-)」 する
+ 「дэг」習慣

(4) 否定形

否定形の語尾は 3 つあり、過去の否定形には、「аагүй」と「сангүй」の語尾を、非過去の否定形では語尾「хгүй」を用いる。前述したように、否定形の語尾「аагүй」は、長母音で始まる語尾であるため、語幹の種類によっては変形が生じる。

語尾「аагүй」や「хгүй」が接続する場合、語尾「v」と同様に、語幹末の ь が母音字 и に変わる。また、語尾「хгүй」では、語幹が子音で終わる場合は、それら語尾の直前に必ず短母音を挟む。

「уншихгүй」読まない
= 「унших (унш-)」 読む
+ 「хгүй」 (и 追加)

3.3 その他の形

(1) 希求形

終止形と同様に、述語として文を終わらすことができる。希求系は、表 2 における意志、願望・希望、依頼・命令が該当し、話し手の希望や意思、聞き手に対する要求などを表す。語尾の変形が生じる一部例を示す。

(a) 命令形

「яваач」行け
= 「явах(яв-)」 行く
+ 「аач」命令

2 人称に対する口語調の文では、語尾をとらず、語幹のみの形をとる場合がある[3]。

(b) 意志

話し手の意志を伝える場合に用いられる。主語が一人称の場合は、「～しましょう」となる。語尾「я (ё, е)」を用いるが、子音で終わる語幹の場合、「ь」または「ъ」の文字が挿入される。

「явья」行こう
= 「явах(яв-)」 行く
+ 「я」 (ъ 追加)

(2) 副動詞形

副動詞形は、主文の述語を修飾する連用節の述語や重文を作る。表 3 に示す語尾を接続するか、または、後置詞をとることによって表される。

(a) 接続助詞

語尾「ч (ж, н)」は、日本語の接続助詞(～して)に相当し、重文の作成時によく現れる。終止形でアスペク

表-3 連用形の語尾の一覧

Table 3 List of particles of the continuative form

| 分類 | モンゴル語 | 日本語 |
|------|--|-----------------|
| 原因 | снаар, снээр, сноор, снө ө р | ～せいで |
| 時間 1 | хад, хэд, ход, хө д ----- ахад, эхэд, оход, ө хө д ахдаа, эхдээ, охдоо, ө хдө ө , ихдаа, ихдоо | ～するとき ～するために |
| 時間 2 | тал, тэл, тол, тө л | ～するまで |
| 時間 3 | саар, сээр, соор, сө ө р | ～してから |
| 目的 | хаар, хээр, хоор, хө ө р | ～するために |
| 接続助詞 | ч, ж, н ----- ач, аж, ан, эч, эж, эн, оч, ож, он, ө ч, ө ж, ө н, ич, иж, ин aad, ээд, оод, ө ө д, гаад, гээд, гоод, гө ө д | ～して |
| 逆説 | вч, ивч, авч, эвч, овч, ө вч | ～しても |
| 仮定 | бал, бэл, бол, бө л, вал, вэл, вол, вө л | ～すれば |

トを表す際にも複合的に用いられる。例えば、現在進行中の動作を表す場合、「ж」語尾の単語に現在時制の補助動詞「байна (いる)」を付け加える。

Захиа бичиж байна.

日本語：手紙を書いている。

「бичиж」書いて
= 「бичих (бич-)」 書く
+ 「ж」～して

語尾「ч (ж, н)」は、次のように文字が追加される場合がある。

「идэн」食べて
= 「идэх (ид-)」 食べる
+ 「н」～して (э 追加)

また、語尾「aad」が用いられることがある。

(b) 仮定形

「орвол」降れば
= 「орох (ор-)」 降る
+ 「вол」 仮定

(c) 時間

① 直ぐに動詞が続く時、次の語尾が使われる。

「идэхэд」食べる時
= 「идэх(ид-)」 食べる
+ 「хэд」～する時

また、この例のように、語尾の先頭に短母音が挿入される場合がある。

② 後に文が続くときは、「ахдаа」という形の語尾となる。

「идэхдээ」食べる時

(d) 後置詞をとる場合

「ирэхээс өмнө」来る前に
 = 「ирэх(ир-)」来る
 + 「хээс өмнө」～する前に

(3) 形動詞形

名詞類を修飾する連体節の述語を作る. 形動詞形も語尾を付ける場合と後置詞を続ける場合がある.

形動詞形の一つに, 動詞を名詞化する名詞節[3]がある. 英語における動名詞であり, 日本語の「～すること」に相当する. モンゴル語では, 形動詞形がそのまま名詞になり, その後に表1の語尾が接続して1つのワードを作る. 時制が現在の形動詞形は動詞の不定形や表2の習慣の語尾が用いられる. 3.2節(1)(c)の体験過去の例で示した不定形「нүүх」がこれに相当する. 時制が過去の場合は, 一般過去が多く用いられる.

習慣や一般過去の語尾に長母音で始まる語尾が接続される際は, 末間母音が脱落してそれぞれ「дг」「сн」に変形する.

(4) 補助語幹

補助語幹[3]は, アスペクトや使役, 受身, 協同などを表すときに用いられる. 語幹の後に補助語幹をはさみ込むことにより, それぞれの形の動詞となる.

使役の語幹 (「уул, лга (лгэ, лго, лгө, лги), га (гэ, го, гө, ги), аа (ээ, оо, өө)」)

受け身の語幹 (「гд, д」)

協同の語幹 (「лц, лд, цгаа (цгээ, цгоо, цгөө)」)

用法の例を以下に示す.

使役 : хүр-эх 届く→ хүр-гэ-х 届かせる

受け身 : хар-ах 見る→ хара-гд-ах 見られる

協同 : ор-ох 入る→ оро-лц-ох 参加する

なお, 受身の語幹「гд」や協同の語幹は, 直前に短母音の追加が生じる場合がある.

よく用いられるアスペクトを表す補助語幹として「чих (完了)」や「схий」[13]がある.

完了 : ид-эх 食べる→ ид-чих-эх 食べてしまう

4. モンゴル語-日本語翻訳支援機能

4.1 機能の概要

処理の概要を図1に示す. 入力されたモンゴル

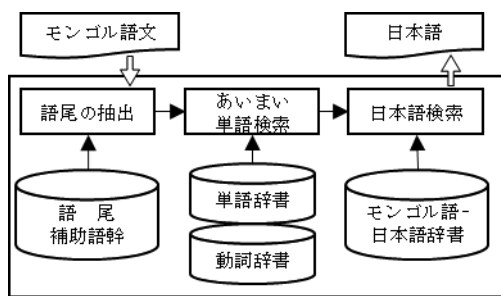


図-1 変換処理の概要

Figure 1 Flow of translation

表-4 モンゴル語-日本語辞書の例

Table 4 Example of mongolian-japanese dictionary

| Mongolian | Japanese | English | Parts |
|-----------|----------|---------|-------|
| авах | 買う | buy | 4 |
| авах | 得る | get | 4 |
| авга | おじ | uncle | 0 |
| авга ах | おじ | uncle | 0 |
| авдар | 箱 | chest | 0 |

語テキストから1行読み込み, スペースやカンマ(,)でワードを1つ1つ切り出す. そして, まず切り出された連続したワード(各ワード間に小文字のスペースを挟んだ文字列)が, 辞書に登録された2単語または3単語の複合語か否かのチェックを行う. 複合語でなければ, 先頭の単独のワードのみの解析を行い, その後, 次のワードを先頭とした複合語のチェックを行う. この処理を繰り返す.

ワード → 語幹 | 不定形 | 名詞 WF | 動詞 WF
 名詞 WF → 語幹 複数形 | 語幹 語尾 |
 語幹 複数形 語尾 | 動詞 WFV 語尾
 動詞 WF → 語幹 語尾 | 語幹 補助語幹 語尾 |
 語幹

ここで, 動詞 WFV は, 動詞 WF のうち 3.3 節(3)の形動詞形の名詞節に示したものである.

単語辞書データベースとして, 次の6つの単語辞書テーブルを作成した. モンゴル語の見出し語(名詞, 形容詞, 副詞の語幹と動詞の不定形など)を格納したモンゴル語単語辞書, モンゴル語の動詞を格納したモンゴル語動詞辞書, 表1の名詞の語尾, 表2と3の動詞の語尾および補助語幹をそれぞれ格納する3つの語尾辞書, モンゴル語と日本語の対応を示した表4のモンゴル語-日本語辞書の6つのテーブルである.

切り出されたワードに対して, 動詞ワードと名詞ワードの両者の場合を想定して処理を行い, 両者の結果を表示する.

(1) まず, 切り出されたワードでモンゴル語単語辞書を検索し, 一致すればモンゴル語-日本語辞書から対応する日本語訳を出力する.

(2) 一致する単語がない場合, まず名詞の解析を行う. 末尾に2.2節(7)の複数形の文字列があれば削除したのち, モンゴル語単語辞書を母音の脱落を考慮したあいまい検索する.

(3) 一致しなければ, ワードの末尾に表1の語尾が付いているか調べ, 語尾を確認したのち削除する. ただし, 表1の語尾の後に再帰語尾が接続されている場合は, 再帰語尾も削除する. さらに, 複数形の文字列があれば削除したのち, モンゴル語単語辞書のあいまい検索を行う. ここでは, 動詞 WFV の検

表-5 モンゴル語動詞辞書の例

Table 5 Example of Mongolian verb dictionary

| Entry word (不定形) | Word stem (語幹) |
|------------------|----------------|
| авах | ав |
| урих | ури |
| урих | урь |

索も行う。

(4) 次に、動詞処理を行う。単語の末尾に表 2 と表 3 の語尾が付いているか調べ、語尾を削除する。

(5) さらに、3.3 節(4)で説明した補助語幹が存在している場合は、補助語幹を削除したのち、モンゴル語動詞辞書から語幹をあいまい検索する。一致すればモンゴル語-日本語辞書を検索し、対応する日本語訳を出力する。文末のワードのときは、命令形を想定して、動詞の語幹を検索する。動詞、名詞ともに該当単語がなければ未知語とする。

4. 2 単語辞書と検索

(1) モンゴル語-日本語辞書

表 4 にモンゴル語単語辞書の内容の一例を示す。辞書には、名詞や副詞、形容詞などの単語の語幹とともに動詞の不定形を登録している。さらに、表 1 に示した方位格の語尾などの後置詞も登録している。また、2.2 節(3)で述べた交代語幹も登録することにした。主格や一部対格は語尾がつかないことがあるため、この辞書とのマッチングにより検出される。なお、表 4 の Parts は、品詞番号である。

この辞書は、モンゴル国立科学技術大学のウブルハンガイ校から提供を受けた英語-モンゴル語電子辞書を基に作成した。本辞典は、英語単語に対するモンゴル語の説明が書かれたもので、この中から辞書として使える部分を取り出し、記載の英単語に基づいて日本語の見出し語の登録を行った。これをベースに公開辞書[3,14-17]を参考に日本語の見出し語の修正や新規単語の追加を行っている。現在、モンゴル語-日本語辞書には、多義を含めて約 20,000 単語を収録している。

(2) モンゴル語単語辞書

モンゴル語-日本語辞書から、モンゴル語見出し語のみを取り出した辞書テーブルである。現在、約 14600 単語収録している。前述したように、辞書には 2 単語以上からなる複合語が約 6600 単語含まれている。複合語は、単語間にスペースを挟んだ形で 1 つの単語として辞書に登録している。また、210 単語の専門用語を登録している。

(3) モンゴル語動詞辞書

モンゴル語単語辞書から動詞だけを取り出したテーブルである。表 5 に内容の一例を示す。動詞辞書は、属性として動詞の見出し語と語幹を持つ。検索時には、辞書の語幹とのマッチングを行い、検出

された単語の見出し語に基づいてモンゴル語-日本語辞書を検索することにより、対応する日本語訳を得ることができる。

ここで、3.2 節(1)(d)で示したように、動詞の語幹が **ь** の文字で終わる場合は、語幹末の **ь** が **и** に変わるという変形規則がある。このため、本来の語幹と変形後の語幹を登録することにした。例えば、似合う「зохих」の場合、本来の語幹「зохь」に変形後の語幹「зохи」を登録した。現在、約 3150 単語収録している。

(4) 語尾および補助語幹

語幹の母音の追加や語尾の変形に対しては、表 1, 2, 3 に示す全ての語尾を追加することにより対応する。ここで、3.3 節で述べたように、動詞の希求形で、語尾「я」が接続するとき、語幹の種類によって硬音符「ь」または軟音符「ъ」を挿入させたうえで、語尾を接続する。このため、「я」のみではなく、「ья」と、「ья」の三種類の語尾を登録し、文字の追加の変形に対応させた。

補助語幹は、3.4 節(4)に示す補助語幹を登録した。ただし、協同は我々にとっては、元の単語に対して意味が大きく異なるため、単語辞書に登録する方針とした。受身、使役についてもよく使われる単語は、できるだけ変化後の単語を単語辞書に登録する。

(5) あいまい検索

2.2 節で述べた母音の脱落に関しては、モンゴル語単語辞書を正規表現による文字列検索を行うことにより対応した。具体的には、切り出された文字列の末尾あるいは末聞母音位置に、1 文字の母音を挿入した正規表現の文字列パターンを作成して検索を行う。例えば、切り出された文字列が「"олн"」であれば、「"ол(а|э|о|ө|и)н"」と「"олн(а|э|о|ө|и)"」の正規表現により、モンゴル語単語辞書の単語を検索する。

4. 3 評価実験

評価プログラムは、Microsoft® Visual Studio® .NET™ 2010 の Visual C++により開発した。また、単語辞書データベースは、Microsoft® Access® を用いて作成した。

モンゴル語のパソコンのアーキテクチャーについて書かれた専門書（日本語訳タイトル「IBM PC のアセンブラ言語」）を用いて有効性の検証を行った。対象は、本テキストの序文を含む 14 ページ、3047 ワードである。

以下に、プログラムの解析結果の例を示す。カギ括弧内（[+〜]）が検出したモンゴル語の語尾で、--の後ろが語尾の日本語訳である。

(原文) Санах ой ба оролт гаралтын төхөөрөмжүүдийг хаяглахад хэрэглэгдэнэ。

表-6 動詞の出現頻度

Table 6 Appearance frequency of verbs

| 分類 | 頻度 | 分類 | 頻度 |
|-------|----|--------|----|
| 過去 | 40 | 原因 | 1 |
| 現在・未来 | 70 | 時間(とき) | 20 |
| 習慣 | 36 | 時間(まで) | 0 |
| 接続助詞 | 72 | 時間(から) | 0 |
| 否定 | 2 | 目的 | 0 |
| 意志 | 5 | 逆説 | 0 |
| 願望・希望 | 0 | 仮定 | 12 |
| 依頼・命令 | 0 | 名詞節 | 2 |

- (1) Санах ой: 記憶, ○記憶装置--
 (2) ба: ○および, ○と, ○そして--
 (3) оролт: 入場, ○入力, 継承--
 (4) гаралтын төхөөрөмж[+ ийг]: ○出力装置 --を
 (5) хаяглах[動]: ○住所を書く
 [助動 + хад--~する時]:
 (名詞) хаяглах[ад]: ○住所を書く--に
 (6) хэрэглэгдэх[動]: ○使う
 хэрэглэгдэх[動]: ○運用する
 хэрэглэгдэх[動]: ○消費する
 хэрэглэгдэх[動]: ○応用する
 хэрэглэгдэх[動]: ○適用する
 хэрэглэгдэх[動]: ○利用する
 [助動 + энэ--(受身)+する]

このフレーズは、パソコンのアドレスバスに関する記述で、「記憶装置と入出力装置のアドレスを指定する(書く)ときに利用される。」との訳になると思われるが、「хаяглах」が「アドレスを書く」という動詞なので、その目的語である「記憶装置」に対格の語尾が接続するものと思われる。このような対訳の扱いは今後の課題である。

上記の様に、各ワードの解析結果では、変形前の単語の見出し語が表示されるため、他の辞書による確認も可能である。また、○のついた日本語訳は、公開辞書等で確認済みの訳であることを示す。(5)のワード「хаяглахад」は、名詞 WF と動詞 WF の両方に判定されている。

3047 ワード中、固有名詞を除く 13 ワード

(0.4%) が未知語とされ解析できなかった。原因の一つに登録していない語尾「ийнх」があった。属格語尾「ийн(の)」と同じであるが、修飾語が被修飾語の後に置かれるときは、この語尾が使われるようである。

また、今回の変換処理では、原則、最長一致法を取っているため、動詞仮定形である「дурдвал(述べると)」が、単独の単語「すなわち」と判断された例があった。同様に、「гэрээс」(家から)が、単独の単語「遺言」に変換された。また、辞

書には動詞の一般過去が形容詞として多く登録されていたため、これに該当する動詞が形容詞と判定されてしまった。文末のワードでは、動詞の解析を優先することにより、平叙文に限れば解決可能であるが、今後の課題である。

名詞 WF と動詞 WF の両方に判定された別の例として、「босоод」が、動詞「бос-оод(起きて)」と名詞「босоо-д(垂直に)」に、「хүртлээ」が、動詞過去「хүрт-лээ(得た)」と名詞「хүртл-ээ(～でさえ/まで)」に判定される例があった。

表 6 に動詞の分類別の出現頻度を示す。序文やテキスト前半のパソコンの歴史に関する記述において、過去形や接続助詞が多かったが、解説が始まると現在形が増えている。全体的に、接続助詞や習慣の語尾が多く用いられることが分かった。過去、現在、習慣、接続助詞のみで全体の 8 割を占める。過去形においては一般過去がほとんどで、体験が 1 例、伝聞が 2 例のみであった。

また、複数形も多くみられ、программууд(プログラム программ), хэлүүд(言語 хэл), битүүд(ビット бит)など欧米圏の言語と同様、単数と複数を使い分けている。

5. むすび

本論文では、人がモンゴル語文章を日本語に翻訳するのを支援する機能の検討を行い、そのために必要となる単語辞書や変換法を提案した。本論文では、モンゴル語から日本語への翻訳に特化することにより、文法上の厳密性には多少欠けるが、母音調和に伴い語尾が複数の形に変化する問題や母音が脱落あるいは挿入される問題を解決し、比較的簡単な処理で原形の単語を検索可能であることを示した。そして評価実験を行った結果、言い回しなどがある程度限られる技術系などの文章であれば十分有用であると思われる。

現在、本評価プログラムを Perl に移植し、Web 環境において 10 ワード程度の文なら数秒で処理可能であることを確認できた。この Web 環境でのツールの利用は、研究や学習の効率を非常に向上させることができると感じた。例えば、文献[3]の例文や、文献[14]に記載された英語の単語や例文を説明した 5 単語前後からなるモンゴル語の文の変換結果は、複合語やモンゴル語の意味や使い方を確認する上で非常に有効であった。そして、変換結果から語尾の使い方や対格の語尾のゼロ語尾化、母音の脱落挿入、母音調和などについて実感することができ、モンゴル語学習のツールとしても有効であると考えられる。

また、モンゴル語に限ったことではないが、本検

討を通して、学習や翻訳支援のためには、複合語が重要であることが確認できた。名詞や動詞のみでなく、例えば、副詞「юу ч болсон (とにかく)」や語尾の言い回し「юу ч биш (何でもない)」などの変換が不可欠である。本ツールでは、複合語を辞書に登録することにより実現できる。

今後の大きな課題として辞書の整備がある。本文で述べたように、今回作成した辞書の日本語見出しは、英語見出しをもとに登録したため、モンゴル語単語の日本語訳として適切な訳になっていないことも少なくなく、変換結果からうまく翻訳できない大きな原因の一つになっている。公開されている辞書[3,15-17]などで確認済みの単語が、現在約9000単語で、半分以上の日本語見出しが確認できていない。今後、単語の新規登録とともに、日本語見出しの見直しも含めて整備を検討していく必要がある。

もう一点は解析結果や日本語訳の候補の削減である。モンゴル語-日本語間にも多くの多義が存在するが、隣接単語の関係だけでも候補を削減できる場合もある。また、モンゴル語の助詞の対格は、日本語の「を」に対応するが、動詞「勝つ」の場合は「に」になるような例外がある[3]。今後、構文解析処理を組み込むことにより、曖昧性の削減や助詞の正しい変換を実現していく予定である。

謝辞

英語-モンゴル語電子辞典や評価用テキストを提供していただいたモンゴル国立科学技術大学のウブルハンガイ校に感謝致します。また、本研究の立ち上げに尽力いただいた Nurul Sakinah Binti Kamaruddin さん、辞書の整備やテキスト入力などに協力いただいた TUMENBAYAR BYAMBADÖRJ さん、MUNKHTSÖG MUNKHZUL さんには心から感謝致します。また、本研究を進めるにあたっては、参考文献 3 から文法や正書法に関して多くの知見を得ました。提供に対して心から感謝致します。

参考文献

- 1) 内田孝: モンゴル語学習者に対する効果的な各語尾の指導法研究, NEAR Conference Proceedings Working Papers, NEAR-2000-10, 国際大学, pp.1-18 (2010).
- 2) Atro Voutilainen, Tanja Purtonen and Kristiina Muhonen: FinnTreeBank2 Manual, University of Helsinki, Department of Modern Languages (2012).
<http://www.ling.helsinki.fi/kieliteknologia/tutkimus/treebank/sources/FinnTreeBankManual.pdf>

- 3) TUFS 言語モジュール (東京外国語大学 21 世紀 COE プログラム「言語運用を基盤とする言語情報学拠点」) モンゴル語・文法モジュール, 東京外国語大学,
<http://www.coelang.tufs.ac.jp/mt/mn/gmod/courses/c02/>
- 4) Mohmoud El-Haj, Udo Kruschwitz and Chris Fox: Creating language resources for under-resourced languages: methodologies, and experiments with Arabic, Language Resources and Evaluation, vol.49, Issue 3, pp. 549-580 (2015).
- 5) 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 回年次大会論文集, pp.709-712 (2004).
- 6) 江原暉将, 早田清冷, 木村展幸: 茶筌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 回年次大会論文集, pp.534-537 (2005).
- 7) Enkhbaya Sanduijav, 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 2004(2004-NL-162), pp.87-93 (2004).
- 8) 大平栄二, 上谷恵里奈, Nurul Sakinah Binti Kamaruddin: モンゴル語-日本語変換に関する基礎的検討, 津山工業高等専門学校紀要, 第 56 号, pp.65-70 (2014).
<http://www.tsuyama-ct.ac.jp/honkou/kyoukan/kiyou/pdf/kiyou2014k02.pdf>
- 9) 河本美月, 大平栄二, 上谷恵里奈: モンゴル語-日本語翻訳支援ツールに関する検討, 平成 27 年度電気・情報関連学会中国支部連合大会講演論文集 (2015).
- 10) 竹嵩志起, Sarenqimuge, 松本忠博: 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, 言語処理学会第 18 回年次大会(NLP2012) 発表論文集 P3-38, pp.1220-1223 (2012).
- 11) モンゴル語正書法,
http://homepage3.nifty.com/itako/mon_jiten019.html
- 12) アルタイ, <http://ameblo.jp/kharkhorin/>
- 13) MANDAKH OYUNCHIMEG: 英語及びモンゴル語の時制、アスペクトに関する対照研究, 京都大学, p.28 (2015).
<http://repository.kulib.kyoto-u.ac.jp/dspace/bitstream/2433/199382/2/dnink00711.pdf>
- 14) Mongolian English Dictionary, <http://www.bolor-toli.com/>
- 15) モンゴル語実践会話入門,
<http://www.aa.tufs.ac.jp/documents/training/ilc/textbooks/2008/mongolian.pdf>
- 16) 橋本勝, エルデネ.プレブジャブ: 現代日本語モンゴル語辞典, 春風社 (2001).
- 17) 長岡技術科学大学工学辞典編纂委員会: 日本語モンゴル語工学用語辞典, 春風社 (2011).

Study on a support tool for Mongolian-Japanese translation

Eiji OHIRA, Erina JYOTANI and Mizuki KOMOTO

Mongolian and Japanese are classified as agglutinative languages. Moreover they have the same sentence structure called SOV. However, some combinations of Mongolian word stems and particles have letter addition or elimination. Mongolian also has many types of vowels and vowel harmonies. These features of Mongolian make it hard to learn Mongolian and to translate into Japanese.

This paper reports on a support tool for Mongolian-Japanese translation and proposes simple rules of grammar. We made an electronic dictionary based on these rules and evaluated the tool experimentally.