

モンゴル語コーパスからの外来語抽出

Khaltar Badam - Osor 藤井 敦 石川 徹也

筑波大学大学院図書館情報メディア研究科 〒305-8550 茨城県つくば市春日 1-2
E-mail: {khab23, fujii, ishikawa}@slis.tsukuba.ac.jp

本稿は、キリル文字のモンゴル語コーパスから外来語を抽出する手法を提案する。モンゴル語における外来語の特徴を分析し、外来語抽出の規則を体系化して抽出に利用する。規則で抽出できない外来語は、日本語のカタカナ語と比較し、カタカナ語と音韻的に類似する場合は外来語として抽出する。さらに、抽出した外来語を日本語と対応付けて日蒙対訳辞書を構築する。評価実験によって本手法の有効性を示す。

Extracting loanwords from Mongolian corpora

Badam- Osor Khaltar Atsushi FUJII Tetsuya ISHIKAWA

Graduate School of Library, Information and Media Studies, University of Tsukuba
1- 2 Kasuga Tsukuba, 305-8550, Japan
E-mail: {khab23, fujii, ishikawa}@slis.tsukuba.ac.jp

This paper proposes a method to extract foreign words from Cyrillic Mongolian corpora. We use hand-crafted rules to extract foreign words from Mongolian corpora, for which we analyzed characteristics of foreign words in Cyrillic Mongolian. We also extract words in Mongolian corpora that are phonetically similar to Japanese Katakana words as foreign words. Additionally, we correspond the extracted foreign words to Japanese words and produce a Japanese-Mongolian bilingual dictionary. We show the effectiveness of our method experimentally.

1. はじめに

近年、新しい技術や概念を表す新語が次々に出現している。ある言語で生まれた新語は、他の言語において音訳されて外来語として移入されることが多い。

新しい外来語は辞書に登録されていないことが多いため、自然言語処理、情報検索、機械翻訳などの精度を低下させる。そこで、新しい外来語を迅速に辞書へ登録することが重要である。しかし、次々に出現する大量の外来語に対して人手で辞書を編集することは時間的、経済的コストがかかる。

この問題を解決するために、コーパスから外来語を自動的に抽出し、辞書を更新する研究が行われている。

モンゴル語においても外来語が頻繁に出現する。外来語をカタカナで表記する日本語とは異なると、モンゴル語では一般語も外来語もキリル文字で表記される。

モンゴル語では、「к」、「п」、「ф」、「ш」という子音は外来語だけに使用される。そこで、これらの子音が使用された単語を「外来語」として一般語から区別することができる。しかし、これらの子音を含まない外来語もある。そのため、モンゴル語のコーパスから外来語を自動的にかつ網羅的に抽出することは容易ではない。

外来語の自動抽出には別の問題もある。モンゴル語の文は句の単位で分かち書きされる。しかし、句は自立語に付属語が接続して構成されており、その結果、自立語と付属語に語形変化が生じるこ

とがある。そこで、外来語を抽出するためには、「語尾処理」によって付属語を分解する必要がある。

本研究は、キリル文字のモンゴル語コーパスから外来語を自動抽出する手法および外来語に対する日本語訳を自動抽出する手法を提案する。

本手法は規則に基づいて外来語を抽出する。さらに、規則で抽出できない外来語を日本語のカタカナ語を手がかりにして抽出する。抽出された外来語と日本語のカタカナ語を比較し、音韻的に類似するカタカナ語を外来語の対訳として抽出する。また、外来語抽出において必要になるモンゴル語の語尾処理手法を提案する。

以下、2章では外来語や対訳の自動抽出およびモンゴル語の語尾処理に関する先行研究について検討する。3章では、本研究で提案する手法を説明し、4章で評価実験について説明する。

2. 先行研究の検討と本研究の位置付け

モンゴル語を対象とした外来語の自動抽出に関する先行研究はない。そこで、モンゴル語以外の言語を対象とした研究について検討する。

Myaeng ら[1] と Jong-Hooh ら[2]は外来語と韓国固有の語との音韻的差異を分析して、外来語であるか韓国語であるかを判断した。しかし、これらの研究では韓国語に特有の処理を行っているため、そのままではモンゴル語に応用することができない。

対訳コーパスから訳語を自動的に抽出することで外来語を抽出する研究[3, 4, 5, 6, 7]がある。しかし、新語の発生に追従して最新のコーパスを利用することができなければ、これらの手法を実用化することはできない。

Nagata ら[8]は、Web から対訳関係にあるページや対訳情報を掲載したページから定期的または安価に対訳を抽出した。しかし、Web には低品質な情報が混在していることがある。また、言語によっては対訳関係にあるページが少ないなどの問題点がある。

金ら[9]は、対訳コーパスを必要とせず、外来語を自動抽出し、更に対訳も自動抽出する手法を提案した。本研究では、金らの研究手法を応用する。しかし、韓国語に特有の処理をモンゴル語に適用させる必要がある。また、処理時間を短縮する必要がある。

モンゴル語の語尾処理に関する既存の手法[10,12]は名詞辞書を利用している。しかし、本研究の目的は辞書に登録されていない外来語を抽出するため、これらの手法を利用することはできない。そこで、名詞辞書を利用しない語尾処理手法を提案する。

3. 本研究で提案する手法

本研究は、モンゴル語コーパスから外来語を抽出し、さらに日本語と対応付けることで日蒙対訳辞書を構築する。システムの構成を図1に示す。点線で囲まれた部分が金ら[9]の手法と異なる処理である。以下、図1の各処理について3.1~3.6で個別に説明する。

3.1 語尾処理

モンゴル語の文では、自立語に付属語が接続して句を構成し、句の単位で分かち書きされる。そこで、空白を用いて機械的に句を抽出することができる。

自立語と付属語が接続するときに語形変化が生じる場合がある。自立語と付属語の接続パターンを図2を用いて説明する。(a)では、「**ном** (本)」に付属語「**ын** (の)」が語形変化なしに接続している。それに対して、(b)~(d)では語形変化が生じている。

(b)では、「**ажил** (仕事)」に付属語「**ын** (の)」が接続する際、下線で示した文字が削除される。(c)では、「**ах** (兄)」に付属語「**д** (に)」が接続する際、下線で示した文字が挿入される。(d)では、「**сургууль** (学校)」に付属語「**аас** (から)」が接続する際、「**сургууль** (学校)」の最後の「**ь**」が削除され、さらに付属語の先頭文字が「**и**」に変化する。

自立語が外来語の場合も付属語が接続し、語形変化が生じることがある。そこで、モンゴル語コーパスから外来語を抽出するためには、外来語に付属語が接続した場合に語尾処理を行い、語形変化を復元しなければならない。外来語は固有名詞や専門用語などの名詞が多いため、名詞句に対して「語尾処理」を行う。金らの研究では語尾処理は韓国語に特有であったので、これをモンゴル語に適用させる。

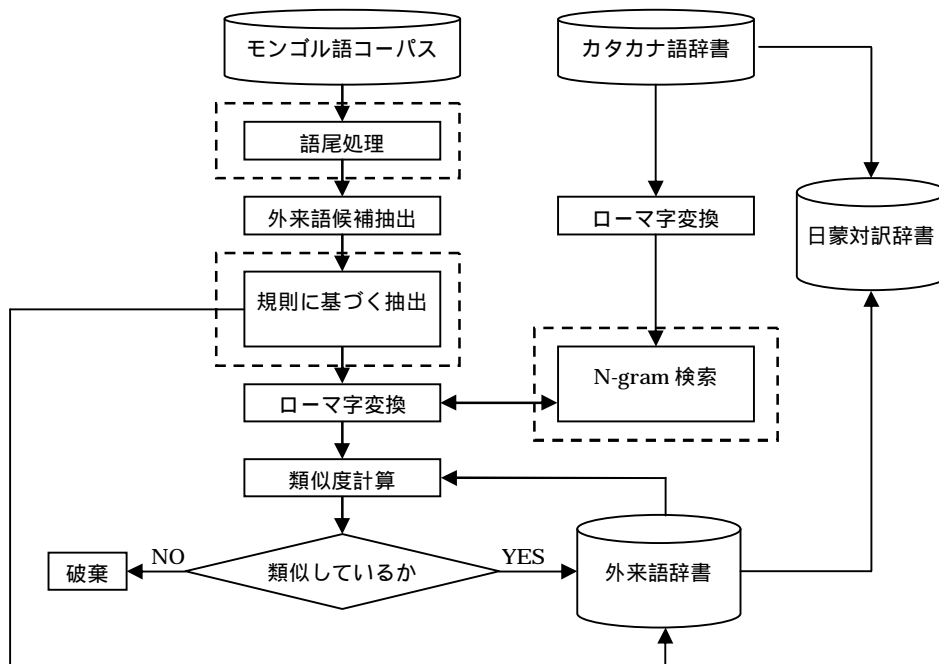


図1 外来語抽出および日蒙対訳辞書構築のシステム構成

パターン	例
(a) 語形変化なしに接続する	НОМ + ЫН → НОМЫН 本 の 本の
(b) 母音の削除	ажил + ЫН → ажлын 仕事 の 仕事の
(c) 母音の挿入	ах + д → ахад 兄 に 兄に
(d) 記号文字 ь が削除され、付属語の母音が и に変化	сургууль + аас → сургуулиас 学校 から 学校から

図2 モンゴル語における自立語と付属語の接続

本研究で提案する語尾処理の流れを図3に示す。まず、入力された単語を「語尾辞書」と後方一致で照合して語尾を検出する。次に「語尾分解規則」を利用し、語尾を分解して名詞を抽出する。さらに、名詞の末尾にある2文字を検査する。名詞の末尾にある2文字が子音の連続であれば、「削除母音の復元規則」を利用して、削除された母音を復元する。そして、母音を復元した名詞を出力する。名詞の末尾にある2文字が子音と母音

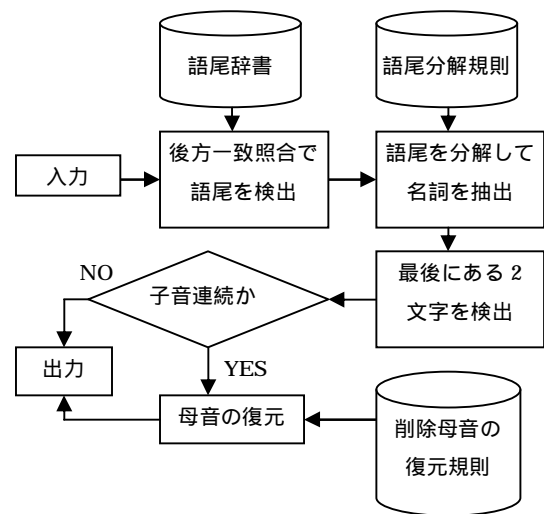


図3 モンゴル語の名詞句に対する語尾処理

の組み合わせであれば、そのまま出力する。このように名詞辞書を使用せずに語尾処理を行うことが本手法の特長である。以下、「語尾辞書」、「語尾分解規則」、「削除母音の復元規則」についてそれぞれ説明する。

語尾辞書

語尾辞書には名詞に接続する語尾が登録されている。語尾辞書の内容を表1に示す。モンゴル語の名詞に接続する語尾は助詞である。ただし、

ある助詞の機能に相当する語尾が複数存在する。

また、名詞に語尾が接続する時に、図 2(d)のように語尾が語形変化する場合があるため、語尾の語形変化後の形態も登録されている。表 1 では括弧の中に語形変化後の語尾を示す。

表 1 モンゴル語の語尾辞書

助詞	名詞に接続する語尾 (括弧内は語形変化後の語尾)
属格(の)	н, ы, ын, ны, ий, ийн, ний
対格(を)	ыг, ийг, г
与位格(に)	д, т
奪格(から)	аас(иас), оос(иос), ээс, өөс
造格(で)	аар(иар), оор(иор), ээр, өөр
共同格(と)	тай, той, тэй
再帰所属	аа(иа), оо(ио), ээ, өө
複数形	ууд(иуд), үүд(иүд)

語尾分解規則

語尾分解規則は、名詞と語尾が接続する時の語尾変化を考慮して作成した。同一の語尾であっても名詞句によって分解する語尾の部分が異なる。本システムでは 173 個の語尾分解規則が登録されている。図 4 に例を示す。

表 1 にある属格の語尾「ийн」は、(a)と(b)どちらの名詞句末尾とも一致する。しかし、一致した語尾を名詞句から分解する時、規則が異なる。名詞句の下線で示した部分が分解される。

(b)では、接続の際に「г」文字が挿入されているため、挿入された部分を削除しなければならない。そこで、一致した「ийн」とともに前の文字も一緒に分解する。

語尾種類	名詞句	名詞
ийн(の)	(a) <u>ээжийн</u> (母の)	ээж(母)
	(b) <u>Хараагийн</u> (ハラー(川の名)の)	Хараа (ハラー)

図 4 語尾分解の例

削除母音の復元規則

語形変化で母音の削除があった場合、名詞を抽出するために削除母音を復元しなければならない。母音の削除があったかどうかは、語尾を分解した後で名詞の語尾にある 2 文字を調べることで分かる。

で分かる。

抽出された名詞の末尾にある 2 文字が子音の連続であった場合、その子音の間にある母音が削除されたと見なす。しかし、元々の名詞が子音の連続で終わる場合もある。そこで、どのような子音連続の時に削除母音を復元するかの 6 通りの規則をモンゴル語の文法教科書[11]を参照して作成した。

例えば、抽出された名詞の最後にある 2 文字は子音「м」, 「г」, 「л」, 「б」, 「в」, 「р」のいずれか 2 つであればそれらの間に母音を復元する。しかし、これらいずれかの後ろに子音「ц」, 「ж」, 「з」, 「с」, 「д」, 「т」, 「ш」, 「ч」, 「х」のいずれかが連続していれば、その間に母音を復元しない。

母音の復元は母音調和規則に従って行う。母音調和規則とは単語の第 1 音節にある母音によってそれ以降にある音節の母音が決まる規則である[11]。例えば、第 1 音節に「а」, 「у」があれば、それ以降の母音は「а」になる。

3.2 外来語候補抽出

モンゴル語コーパスから外来語ではない単語を除去して、外来語候補を絞る。

具体的には、既存のモンゴル語辞書に登録されている語を削除する。ここでは、名詞辞書は Enhbayar ら[10]が作成した辞書である。この辞書には名詞が 1926 語が登録されている。

また、「規則に基づく抽出」で誤って抽出されるモンゴル語の人名などの固有名詞と略語も削除する。モンゴル語文において、固有名詞の先頭文字は大文字で書かれている。略語は全て大文字で書かれている。そこで、コーパスにおいて文頭以外の場所で大文字ではじまる単語を削除する。また、西洋言語に使用されない「ө」, 「ү」母音を含む語も削除する。

3.3 規則に基づく抽出

モンゴル語コーパスから外来語を規則に基づいて抽出する。以下の規則(a)~(g)のいずれかに当てはまる単語を外来語候補として抽出する。

- (a) モンゴル語における外来語に特有な 4 つの子音「к」, 「п」, 「ф」, 「ш」のいずれかを含む単語

- (b) 母音調和規則に違反する単語
モンゴル語では、人名などの固有名詞以外は男性母音と女性母音が同一の単語に混在しない。そのため、男性母音と女性母音が混在している単語は外来語である可能性が高い。
- (c) 語頭が子音の連続である単語
モンゴル語において、語頭は子音の連続で始まらないため、語頭は子音の連続である単語は外来語である可能性が高い。
- (d) 語尾が特定の子音連続である単語
モンゴル語の子音「 Π 」, 「 δ 」, 「 T 」, 「 Π 」, 「 Ψ 」, 「 Σ 」, 「 Π 」の後ろに子音が入る時には必ず母音を挟む[13]。この規則に違反する単語は外来語である可能性が高い。
- (e) 「 B 」で始まる単語
子音「 B 」が語頭に入るのは特定の単語だけである。現代モンゴル語辞典[14]には、「 B 」で始まる語が 54 あった。その中で、西洋言語からの外来語は 31 語であり、現在使われているモンゴル語は 8 語であり、残りの 15 語はほとんど使われない単語であった。そのため、「 B 」で始まる単語は外来語である可能性が高い。
- (f) 「 p 」で始まる単語
現代モンゴル語辞典[14]では、「 p 」で始まる単語は 49 あり、そのうち 4 語がモンゴル語固有の語であった。そのため、「 p 」で始まる単語は外来語である可能性が高い。
- (g) 語尾が「子音+и」である単語

3.4 ローマ字変換

同一言語から移入した外来語は音韻的に似ているという特徴がある。英語から移入されたモンゴル語と日本語の外来語は音韻的に類似している。またローマ字表記も似ている。英語の「system」から移入した日本語の「システム」

とモンゴル語の「систем」のローマ字表記はそれぞれ「shisutemu」と「sistem」であり類似している。

そこで、規則で抽出できない外来語は、日本語のカタカナ語を手がかりにして抽出する。具体的に、日本語のカタカナ語と比較して、類似するモンゴル語の単語を外来語として抽出する。

日本語とモンゴル語を比較するために、まず両言語を同等に比較できる中間言語に変換する。中間言語としてローマ字を使用する。

ローマ字表記の時に、前処理でモンゴル語の 12 通りの表記を日本語ローマ字に合わせる。例えば、モンゴル語の子音「 Π 」はローマ字で「l」と表記する。しかし、日本語のローマ字表記には「l」を使用しないので、日本語に合わせて「r」に統一する。

3.5 N-gram 検索

金らの手法[9]ではカタカナ語と韓国語をローマ字に変換し、DP マッチングを用いて比較するため処理時間が遅い。そこで、本研究では N-gram 検索を用いてこの問題を解決する。

N-gram 検索は文字列の順番を考慮しないため類似度計算の精度が低い反面、処理時間が速い。これに対して DP マッチングは文字列の順番を考慮するため類似度計算の精度が高い反面、処理時間が遅い。本研究では両手法の利点を使う。すなわち、外来語候補を検索質問として、N-gram 検索によってカタカナ語辞書から類似するカタカナ語の候補を絞る。次に、各候補に対して DP マッチングを適用し、類似度計算の精度を上げる。

N-gram 検索は 2 つの文字列を N 文字単位で比較する。本研究では $N=2$ とする。また、検索モデルとして Okapi BM25 [15]を使用する。

3.6 類似度計算

N-gram の検索質問として使用した外来語候補と検索された各カタカナ語に対して DP マッチングで類似度を計算する。そして、カタカナ語と類似度が高い単語を外来語として抽出する。

DP マッチングはパターンの変異を「挿入」, 「削除」, 「置き換え」の 3 種類で評価し、2 つの文字列の類似度を最小の差異数で測定する。

同一言語からきたカタカナ語とモンゴル語外来語のローマ字表記において、子音があまり変わらないことに対して、各言語の体系によって母音

の揺れがある。そこで、子音を重視して重みを加える。カタカナ語とモンゴル語外来語候補の類似度は、式(1)によって計算する。

$$\text{類似度} = 1 - \frac{2 \times (w \times \text{子音差異数} + \text{母音差異数})}{w \times \text{子音の数} + \text{母音の数}} \quad (1)$$

類似度は0~1の値をとる。wは子音の重要度を制御するパラメータであり、経験的にw=2としている。

ある閾値以上の類似度を持つカタカナ語とモンゴル語外来語候補の対を抽出して、モンゴル語の外来語として外来語辞書に登録する。更に、日本語のカタカナ語とモンゴル語外来語を対訳として日蒙対訳辞書に登録する。

なお、規則に基づく抽出(3.3節)で抽出された外来語についても、3.5節と3.6節の手法を用いて、日本語のカタカナ語と対応付けて対訳辞書に登録する。

4. 評価実験

本システムの有効性を評価するために、専門用語辞書から抽出したカタカナ語辞書を利用した。この辞書は111,166語を含んでいる。モンゴル語コーパスとして、「Mongolian IT Park」¹のWebサイトから収集した研究抄録1,118件(1994~2004年)を用いた。

本研究で提案したシステムを「語尾処理」、「外来語抽出」、「日蒙対訳抽出」、「処理時間」の観点から評価した。以下、4.1~4.4節でそれぞれの評価について説明する。

4.1 語尾処理の評価

モンゴル語コーパスの中から10文書を無作為に選び、語尾処理の評価に使用した。分野の内訳は「農業(5)」、「社会(3)」、「医学(2)」であった。語尾処理によって名詞を抽出し、精度(正解率)を評価した。

実験の結果を表2に示す。一般名詞に対する語尾処理の精度が99.5%、外来語に対する語尾処理の精度が98.8%という良好な結果を得ることができた。

表2 語尾処理の実験結果

	語数	精度(%)
一般名詞	367	99.5
外来語	82	98.8

抽出に失敗した原因は、削除母音復元における誤り(2件)、「д」で終わる単語を与格の「д」と誤って削除したこと(1件)であった。現在のところ、これらの誤りに対する根本的な解決策はない。

4.2 外来語抽出の評価

モンゴル語コーパスには異なりで27,482語が含まれていた。この中から高頻度語1,300語を選択し、外来語抽出の評価に使用した。1,300語のうち、外来語は165語あった。外来語抽出の精度と再現率は式(2)を用いて計算した。

$$\text{精度} = \frac{\text{システムが出力した正しい外来語数}}{\text{システムが出力した外来語総数}} \quad (2)$$

$$\text{再現率} = \frac{\text{システムが出力した正しい外来語数}}{\text{コーパスに出現した外来語総数}}$$

まず、3.3節で説明した規則に基づいて抽出した結果、139語が外来語として抽出された。表3に規則ごとの精度と再現率を示す。表3における(a)~(g)は3.3節の規則に対応する。

表3の結果より、全規則を利用すると精度が悪くなるものの、再現率は最も良かった。人手で修正するための前処理として本手法を使用する場合は再現率が重要である。このような場合は全規則を使用することが有効である。

次に、規則で抽出されなかった外来語をカタカナ語によって抽出した。類似度の閾値を0.6に設定し、さらに類似度順位を上位5位まで考慮した結果、12語が外来語として抽出された。これらの外来語と規則で抽出された外来語を合わせると、精度と再現率はそれぞれ71.9%と91.5%となった。すなわち、カタカナ語を用いると精度は落ちるものの、再現率を高めることができた。

4.3 日蒙対訳抽出の評価

モンゴル語コーパスから抽出した外来語には、対応するカタカナ語を人手で特定できないものがあつた。そのような外来語は正解判定が難しい

¹ <http://www.itpark.mn/>

表3 外来語抽出の精度と再現率

規則	(a)	(b)	(c)	(d)	(e)	(f)	(g)	全規則
自動的に抽出された語数	102	63	21	6	4	5	24	150
抽出された正しい外来語数	101	60	20	5	4	5	19	139
精度 (%)	99.0	95.2	95.2	83.3	100	100	79.2	92.7
再現率 (%)	61.2	36.4	12.1	3.0	2.4	3.03	11.5	84.2

ため、カタカナ語を特定できる外来語 124 語のみを評価の対象とした。

類似度の閾値を 0.6 に設定して、さらに類似度順位を上位 5 位まで考慮したところ、109 語の対訳が抽出された。対訳抽出の精度は 56.3%、再現率は 87.9%であった。図 5 に自動抽出された対訳のうち正しかったものを示す。

対訳抽出に失敗した原因は、対応するカタカナ語が辞書に登録されていなかったこと(7件)と類似度が閾値の 0.6 より低かったこと(5件)、N-gram 検索で候補として特定されなかったこと(3件)であった。

日本語	モンゴル語
アリブミン	альбумин
ラボラトリー	лаборатор
マネージメント	менежмент
メカニズム	механизм
ミトコンドリア	митохондр
モニターリング	мониторинг
モルフォロジー	морфологи
オペレーター	оператор
パラメーター	параметр
パテント	патент
ポリオウイルス	полиовирус
ポテンシャル	потенциал
プログラム	программ
プロトン	протон
プロセス	процесс
レセピエント	рецепент
セミナー	семинар
ストラテジー	стратегии
テクトニック	тектоник
テクノロジー	технологи
ファーマコロジー	фармакологи
フィジオロジー	физиологи
エコロジー	экологи

図5 自動抽出された日蒙対訳の例

4.4 処理時間の評価

モンゴル語コーパスから外来語 102 語を無作為に選択し、「N-gramのみ」、「DP マッチングのみ」、「N-gram + DP マッチング」の手法で個別に計算を行い、処理時間と正解順位を比較した。実験を行った計算機環境は「CPU : Pentium 1GHz (×2), メモリ : 2GB」であった。

実験結果を表 4 に示す。なお、DP マッチング単体で抽出されたものの、N-gram 単体では抽出されなかった正解外来語は 3 件あった。これらは表 4 には含めていない。

表 4 の結果より、N-gram と DP マッチングを併用することで、DP マッチング単体よりも処理時間を短縮し、N-gram 検索単体よりも正解順位を向上させることができた。

表4 処理時間の比較

手法	N-gram	DP	N-gram+DP
外来語数	102		
処理時間	95 秒	38 時間 15 秒	4 分 53 秒
正解が抽出された語数	66	66	66
正解の平均順位	44.8 位	2.7 位	2.7 位

おわりに

本研究はモンゴル語コーパスから外来語を抽出する手法を提案した。まず、外来語を規則に基づいて抽出した。次に、規則で抽出できない外来語を日本語のカタカナ語を手がかりにして抽出した。その際に、名詞辞書に依存しない語尾処理の手法を提案した。また日蒙対訳辞書を構築する手法を提案した。システムの処理時間も先行研究より改善された。

今後の研究課題として、日本語とモンゴル語における外来語表記の特徴を分析して外来語や対訳の抽出精度を向上させる必要がある。

謝辞

本研究では Sanduijav Enkhbayar 氏が作成した名詞辞書を使用させて頂きました。心から感謝いたします。

参考文献

[1] Sung Hyun Myaeng and Kil-Soon Jeong. Back-Transliteration of Foreign Words for Information Retrieval. *Information Processing and Management*, Vol. 35, No. 4, pp. 523 - 540, 1999.

[2] Jong-Hooh Oh and Key-Sun Choi. Automatic Extraction of Transliterated Foreign Words Using Hidden Markov Model. *International Conference on Computer Processing of Oriental Languages*, pp. 433 - 438, 2001.

[3] 山本由紀雄, 松本仁. 対訳コーパスを用いた専門用語対訳辞書の作成. *情報処理学会研究報告*, NL - 94, pp. 85 - 92, 1993.

[4] 石本浩之, 長尾真. 対訳文章を利用した専門用語対訳辞書の自動作成: 訳語対応における両立不可能性を考慮した手法について. *情報処理学会報告*, NL - 102, pp. 81 - 88, 1994.

[5] 松尾義博, 白井論. 発音情報を用いた訳語対の自動抽出. *情報処理学会研究報告*, 96 - NL - 116, pp.101 - 106, 1996.

[6] Nigel Collier, Akira Kumano, and Hideki Hirakawa. Acquisition of English- Japanese proper nouns from noisy-parallel newswire articles using Katakana matching. *Proceedings of the Natural Language Processing Pacific Rim Symposium*, pp. 309 - 314, 1997.

[7] Keita Tsuji. Automatic Extraction of Translational Japanese - KATAKANA and English Word Pairs from Bilingual Corpora. *International*

Journal of Computer Processing of Oriental Languages, Vol. 15, No.3, pp. 261 - 279, 2002.

[8] Masaaki Nagata, Teruka Saito, and Kenji Suzuki. Using the Web as a bilingual dictionary. *Proceedings of the ACL - EACL Workshop on Data-Driven Machine Translation*, pp. 95 - 102, 2001.

[9] 金玉錦, 藤井 敦, 石川 徹也. 韓国語コーパスからの外来語自動抽出と言語解析への応用. *言語処理学会第 9 回年次大会発表論文集*, pp. 258 - 261, 2003.

[10] 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形態素解析. *情報処理学会研究報告*, 2004 - NL - 164, pp.41 - 46, 2004.

[11] Ц. Баярмаа. Монгол хэл I - IV анги. 2002. (1年生から4年生のモンゴル語文法)

[12] 江原輝将, 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, 2004.

[13] С. Цэгмид, Ж. Оюунцэцэг. Монгол хэлний хураангуй лавлах. 1997. (モンゴル語の文法書)

[14] 小沢重男. 現代モンゴル語辞典. 大学書林. 2000.

[15] S. E. Robertson, S. Walker, S. Jones, M. M. Hancock-Beaulieu, and M. Gatford. Okapi at TREC - 3, *Proceedings of the Third Text REetrieval Conference (TREC - 3)*, NIST Special Publication 500 - 226, 1995.