伝統的モンゴル語と現代モンゴル語を対象とした双方向的な翻字手法

満 都 拉[†] 藤 井 敦[†] 石 川 徹 也[††]

モンゴル語には，モンゴル文字を用いて表記する伝統的モンゴル語とキリル文字を用いて表記する現代モンゴル語の 2 種類がある．伝統的モンゴル語は主に中国の内モンゴル自治区で使われており，現代モンゴル語は主にモンゴル国で使われている．両方のモンゴル語を読み書きができる人は少ないため，内モンゴル自治区とモンゴル国で情報の交換が困難である．しかし，2 つのモンゴル語は音声言語としてはほとんど同じであり，発音に基づいて文字単位の対応を付けることができる．そこで，本論文は伝統的モンゴル語と現代モンゴル語を双方向的に翻字する手法を提案する．具体的には，一方のモンゴル語で書かれたテキストを文字単位で他方のモンゴル語に変換する．また，正字法を適用し，文字単位では形式化が困難な表記上の違いに対処する．新聞記事を用いた評価実験の結果，現代モンゴル語から伝統的モンゴル語への翻字精度は 80.6%，伝統的モンゴル語から現代モンゴル語への翻字精度は 85.5%であった．また，本手法による自動翻字の結果に誤りが含まれてもテキストの内容理解には支障がなかった．

A Bidirectional Transliteration Method for the Traditional and Modern Mongolian Scripts

Dula Man,[†] Atsushi Fujii[†] and Tetsuya Ishikawa[††]

The Mongolian language is divided into the Traditional Mongolian that is written with the Mongolian alphabet and the Modern Mongolian that is written with the Cyrillic alphabet. The Traditional Mongolian has mainly been used in Inner Mongolian Autonomous Region in China. The Modern Mongolian has mainly been used in Mongolia. Because few people have a good command of both Mongolian languages, information interchange between Inner Mongolian Autonomous Region and Mongolia is difficult. However, because the two Mongolian languages are similar to each other as spoken languages, the alphabets of the two languages can be aligned on the basis of pronunciation. This paper proposes a method to transliterate the Traditional Mongolian script and the Modern Mongolian script bidirectionally. We convert script in one Mongolian language into the other Mongolian language and use rules for orthography to counter the differences in the two languages that cannot easily be formalized on a character-by-character basis. Experiments using newspaper articles showed that the transliteration accuracy from the Modern Mongolian to the Traditional Mongolian was 80.6% and that the transliteration accuracy from the Traditional Mongolian to the Modern Mongolian was 85.5%. Additionally, errors in the automatic transliteration did not decrease the degree of comprehension for the source text.

1. はじめに

モンゴル語には，モンゴル文字で表記する「伝統的モンゴル語」，キリル文字で表記する「現代モンゴル語」，改良モンゴル文字で表記する「トド文字モンゴル語」がある．本研究では伝統的モンゴル語と現代モンゴル語を対象とする．

伝統的モンゴル語は，主に中国の内モンゴル自治区で使われている．諸般の理由によって，現代モンゴル語は内モンゴル自治区で普及しなかった．そこで，内モンゴル自治区では現代モンゴル語の読み書きができる人は少ない．

現代モンゴル語は，主にモンゴル国で使われている．モンゴル国では，1946 年にキリル文字が公式な国字として採用され，伝統的モンゴル語の使用は廃止された．また，1980 年代後半まで中国の内モンゴル自治

† 筑波大学大学院図書館情報メディア研究科
Graduate School of Library, Information and Media Studies, University of Tsukuba
†† 東京大学史料編纂所前近代日本史情報国際センター
International Center for Digitization of Premodern Japanese Sources, the Historiographical Institute, The University of Tokyo

区と交流が少なかった．そこで，モンゴル国では伝統的モンゴル語の読み書きができる人は少ない．

以上の理由から，内モンゴル自治区とモンゴル国の間で言葉（書き言葉）の壁が生じ，情報の交換が困難になっている．

伝統的モンゴル語と現代モンゴル語は，話し言葉としては類似しており，表記に使用する文字体系が根本的に異なる．しかし，両モンゴル語における表記上の差異は，限られた数の規則によっておおむね体系化することが可能である．

本研究は，この性質に着目し，一方のモンゴル語で書かれたテキストを他方のモンゴル語に文字単位で変換するための翻字手法を提案する．その結果，伝統的モンゴル語と現代モンゴル語のテキスト情報を自動的に相互変換することができ，内モンゴル自治区とモンゴル国の間で情報交換が容易になる．

また，伝統的モンゴル語のテキストは電子化が進んでおらず，テキスト処理に関する研究が遅れている．しかし，現代モンゴル語テキストの電子化は進んでいる．そこで，現代モンゴル語の電子化テキストを伝統的モンゴル語に翻字することによって，伝統的モンゴル語のテキスト処理研究を促進することが期待できる．

言語処理において，十分な訓練データがあれば機械学習によって高い精度を達成できる場合がある．しかし，伝統的モンゴル語には十分な量の電子化されたデータがないため，人手で作成した規則に基づく翻字手法を提案する．

以下，2 章で伝統的モンゴル語と現代モンゴル語の相違点について説明し，3 章で関連研究について検討する．4 章で本研究の翻字手法について説明し，5 章で本手法の評価実験について説明する．

2. 伝統的モンゴル語と現代モンゴル語の違い

2.1 概　　要

伝統的モンゴル語は，上から下へ縦書きし，行は左から右へ進む．伝統的モンゴル語のテキスト例を**図 1** に示す．

印刷の都合上，図 1 以外の図表や本文ではモンゴル文字を左に 90° 傾けて表記する．

現代モンゴル語はヨーロッパの諸言語と同様に左から右へ横書きし，行は上から下へ進む．図 1 と同じ内容の現代モンゴル語テキストを**図 2** に示す．

伝統的モンゴル語と現代モンゴル語は，どちらも「母音調和規則」に従う[1)~4)]．この規則は，母音を陽性（男性），中性，陰性（女性）に分け，1 つの単語に陽性母音と陰性母音が混在することを禁止している．

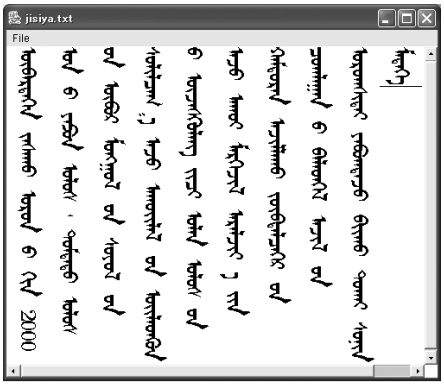


図 1 伝統的モンゴル語テキストの例
Fig. 1 Example text in the Traditional Mongolian.

өөртөө засах орных нь 2000 оны Япон улс , дундад улсын өвөр монголын соёлын солилцоо , аж ахуйллын үйлдэхний үзэсгэлэн жич олон улсын аж ахуй мэргэжил арилжааны хамтран ажиллах зөвлөлцөөрийн цуглааны бэлтгэл ажлын урагштай явагдаж байх тухай сонин мэдээ

図 2 図 1 と同じ内容の現代モンゴル語テキスト
Fig. 2 Example text in the Modern Mongolian representing the same meaning as Fig. 1.

中性母音は陽性母音と陰性母音のどちらとも混在して単語を構成することができる．

陽性母音だけか，陽性母音と中性母音で構成された語を「陽性語」と呼ぶ．陰性母音だけか，陰性母音と中性母音で構成された語を「陰性語」と呼ぶ．

しかし，地名や人名などの固有名詞は母音調和規則に従わず，陽性語と陰性語が混在して複合語を構成する場合がある．これは，外来語でないモンゴル語の固有名詞でも起こる．

2.2 文字体系の違い

2.2.1 伝統的モンゴル語の文字体系

伝統的モンゴル語で使用されるモンゴル文字を**図 3** に示す．モンゴル文字には，基本母音が 7 種類（陽性 3，中性 1，陰性 3），子音が 24 種類ある[2)~4)]．また，2 つの基本母音を連続して発音する「双母音」が 9 種類ある．しかし，モンゴル文字には長母音を表記する文字がない．詳細は 2.3.1 項で説明する．図 3 におい

図 3　伝統的モンゴル語で使用されるモンゴル文字
Fig. 3　The Mongolian alphabet used in the Traditional Mongolian.

基本母音	陽性	a o y
	中性	и
	陰性	э ө ү
補助母音	陽性	я ю ё ы
	中性	й
	陰性	е ю
長母音	陽性	аа уу оо яа ёо юу
	中性	ий
	陰性	ээ өө үү еэ еө юү
双母音	陽性	ай ой уй яй ёй юй
	陰性	эй үй юй ей
子音		б в г д ж з к л м н п р с т ф х ц ч ш щ
音符	軟音符	ь
	硬音符	ъ

図 4　現代モンゴル語で使用されるキリル文字
Fig. 4　The Cyrillic alphabet used in the Modern Mongolian.

て，同じ字形の文字が繰り返し出現する場合がある．具体的には「」，「」，「」，「」である．しかし，字形は同じでも発音が異なるため別の文字として扱う．

なお，本研究では，満ら[6] が開発した手法を用いて伝統的モンゴル語を電子化する．当手法は発音情報に基づいてモンゴル文字を保存するため，字形が同じでも発音が異なる文字は区別される．

2.2.2　現代モンゴル語の文字体系

現代モンゴル語で使用されるキリル文字を**図 4** に示す．キリル文字はロシア語でも使用されている．ただし，ロシア語のキリル文字（33 文字）にモンゴル語特有の発音に対応する「ө」と「ү」が追加され，合計 35 文字ある．その内訳は，基本母音 7，補助母音 6（「ю」は陽性と陰性に共通），子音 20，音符 2 である[1]．音符とは，子音の音韻変化（軟化）や子音と母音の音韻分離を行うための記号である．直前にある子音の音韻を軟化する役割を持つ音符を「軟音符」と呼び，子音と母音の音韻を分離する役割を持つ音符を「硬音符」と呼ぶ．ただし，音符自体は発音しない．

また，基本母音を組み合わせて表記する長母音が 13 種類，双母音が 9 種類ある（юйは陽性と陰性に共通）．

2.3　単語表記の違い

伝統的モンゴル語と現代モンゴル語の単語表記における違いは，a）助詞の分かち書き，b）長母音の表記法，c）正字法の 3 点である．正字法とは，語を構成する子音と母音の接続に関する規則である．2.3.1 と 2.3.2 項で各モンゴル語における単語表記について説明する．

2.3.1　伝統的モンゴル語の単語表記

伝統的モンゴル語の文字は，語中の位置によって字形が異なる．語頭における字形を「語頭形」，語末尾における字形を「語尾形」，語頭と語尾以外の位置における字形を「語中形」と呼ぶ．1 文字が独立して書かれる字形を「独立形」と呼ぶ．たとえば，母音の「a」は独立形「」，語頭形「」，語中形「」，語尾形「」もしくは「」の 5 種類の字形を持つ．

また，伝統的モンゴル語では助詞を分かち書きする場合がある．分かち書きされる助詞の先頭文字は語頭形ではなく語中形をとり，語の性によって字形は変わらない．しかし，直前にある自立語の末尾が子音か母音かによって字形が異なる．子音で終わる語の後ろには母音で始まる助詞を使い，母音で終わる語の後ろには子音で始まる助詞を使う．よって，母音で始まる助詞と子音で始まる助詞の 2 種類がある．

伝統的モンゴル語の正字法では形態論を重視するため，現在は発音しなくなった母音（弱化母音）も表記する．たとえば，日本語の「位置」の意味を表す語は，伝統的モンゴル語で「」（bai-ri）のように 2 音節で表記する．しかし，実際に発音するときは「bair」のように 1 音節で発音する．末尾の音節「」（ri）の母音「i」は弱化母音であるため発音しない．

また，伝統的モンゴル語には長母音を表記する文字がないため，以下に示す 4 種類の方法で長母音を表記する．

- 音節による長母音化

　子音「g」と母音が結合した音節がよく使われ

る．たとえば，日本語の「赤」を伝統的モンゴル語で「ᠤᠯᠠᠭᠠᠨ」（olaan）と表記する．ここでは音節「ᠭᠠ」（ga）を用いて直前の母音「ᠠ」（a）が長母音化され，「aa」と発音する．

- 双母音による長母音化
  たとえば，日本語の「子供」をモンゴル語で「ᠦᠭᠦᠬᠡᠳ」（huuhed）と表記する．ここでは，母音「ᠡ」（e）と「ᠦ」（u）で構成された双母音「ᠡᠦ」（eu）で長母音（uu）が表記されている．
- 同じ母音の繰返しによる長母音化
  たとえば，日本語の「息子」をモンゴル語で「ᠦᠦ」（huu）と表記する．ここでは，同じ母音の繰返し「ᠦᠦ」（uu）で長母音「uu」を表記している．
- 短母音による長母音化
  たとえば，日本語の「父親」をモンゴル語では「ᠠᠪᠤ」（abo）と表記する．語頭の短母音「ᠠ」（a）を長母音の（aa）のように発音して，「ᠠᠠᠪ」を（aab）と発音する．

2.3.2 現代モンゴル語の単語表記

現代モンゴル語では助詞を分かち書きしない．助詞は活用語尾として自立語に接続される．そこで，母音調和規則に従って，接続される自立語の性により活用語尾の母音が異なる．すなわち，陽性語には陽性母音の活用語尾が接続されて，陰性語には陰性母音の活用語尾が接続される．

また，現代モンゴル語では長母音を表記する文字がある．しかし，現代モンゴル語では弱化母音を表記しないため，伝統的モンゴル語で弱化母音によって区別される同音異議語を現代モンゴル語では区別できない場合がある．たとえば，日本語の「愛」と「砂利」の意味を表す 2 つの語は，モンゴル語ではどちらも「hair」と発音する．伝統的モンゴル語では「愛」を「ᠬᠠᠢᠷᠠᠨ」と表記し，「砂利」を「ᠬᠠᠢᠷᠠ」と表記する．ここで，末尾の弱化母音「ᠨ」によって「愛」と「砂利」の意味が区別される．しかし，現代モンゴル語では両方とも「xaйp」と表記するため，文脈がなければ「愛」と「砂利」の意味を区別することができない．

3. 関連研究と本研究の違い

中里ら[5]は，伝統的モンゴル語から現代モンゴル語へ翻字するために最適な変換単位と正字法の規則化について検討した．

最適な変換単位の判断基準として，同じ単語をそれぞれ伝統的モンゴル語と現代モンゴル語で表記し，文字や音節などの単位で分割したときに，両モンゴル語の分割数が等しくなることを条件とした．その結果，母音に挟まれる「g」を考慮した分割方法が比較的良かったものの，例外が多くて実用的ではない．また，彼らは変換単位や正字法の規則化について検討しただけで翻字手法を実現するに至っていない．

中里らは，伝統的モンゴル語から現代モンゴル語への変換について検討した．しかし，現代モンゴル語から伝統的モンゴル語への変換については検討していない．

また，中里らの変換対象は単語であり，文章（テキスト）ではない．伝統的モンゴル語と現代モンゴル語は助詞の表記法が異なるため，文章を対象とする場合は文脈に応じた助詞や活用語尾の処理が必要である．すなわち，中里らの手法を単語単位に適用しても文章を正しく翻字することはできない．

そこで，本研究は文章を変換の対象として，伝統的モンゴル語と現代モンゴル語の双方向的な翻字手法を提案する．

4. 本研究で提案するモンゴル語の翻字手法

4.1 概　　要

本研究で提案する伝統的モンゴル語と現代モンゴル語の翻字手法を**図 5** に示す．以下，図 5 を用いて処理の概要を説明する．

まず，「単語の抽出」によって，原言語テキストから単語を抽出する．次に，「助詞の処理」を行う．ただし，処理の内容は原言語がどちらのモンゴル語かによって異なる．

原言語が伝統的モンゴル語の場合は，抽出された単語が助詞かどうかを調べて，助詞の場合は直前の自立語に接続する．

原言語が現代モンゴル語の場合は，抽出された単語に活用語尾が含まれており，かつその語尾が伝統的モンゴル語の助詞に相当するかどうかを調べる．助詞に

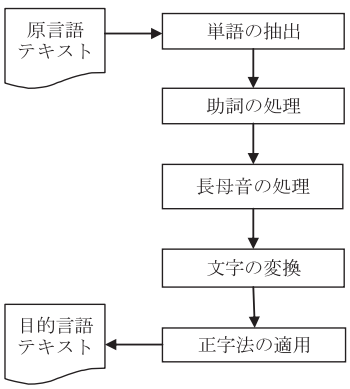


図 5 本研究で提案するモンゴル語翻字手法の概要
Fig. 5 Overview of the transliteration method proposed in this paper.

相当している場合は活用語尾を分割する．

「長母音の処理」では，伝統的モンゴル語と現代モンゴル語の長母音に関する差異に対処する．

「文字の変換」では，原言語テキストを文字単位で目的言語に変換する．しかし，伝統的モンゴル語と現代モンゴル語の正字法が異なるため，文字単位で変換しただけでは目的言語の正字法に従わない語が含まれることがある．そこで，「正字法の適用」によって，文字変換の結果を修正し，目的言語テキストを出力する．

以下，4.2 節で現代モンゴル語から伝統的モンゴル語への翻字について説明し，4.3 節で伝統的モンゴル語から現代モンゴル語への翻字について説明する．

4.2 現代モンゴル語から伝統的モンゴル語への翻字

現代モンゴル語では，助詞を分かち書きせずに活用語尾として前の自立語に接続する．しかし，伝統的モンゴル語では助詞を分かち書きする．そこで，現代モンゴル語テキスト中の活用語尾を助詞として分割する必要がある．また，長母音を伝統的モンゴル語の表記（2.3.1 項参照）に変換する必要がある．以下，図 5 の手順に従って説明する．

4.2.1 単語の抽出

現代モンゴル語のテキストは分かち書きされるため，空白によって単語を抽出する．ただし，助詞は直前の自立語に接続されて文節を構成している．

4.2.2 助詞の処理

伝統的モンゴル語の助詞（格助詞，二重格，再帰格など）は分かち書きされる．しかし，現代モンゴル語の助詞は自立語に語尾として接続されている．そこで，現代モンゴル語の語尾を分割し，さらにそれを伝統的モンゴル語の助詞に変換する規則を作成した（**図 6** と **図 7**）．ただし，図 6 は二重格の助詞に関する規則である．単語抽出によって抽出された現代モンゴル語の単語に図 6 と図 7 の語尾が含まれている場合，その語幹末尾の文字を参考にして，語尾を伝統的モンゴル語の助詞に変換する．図 6 と図 7 の語尾が含まれない場合は自立語と見なす．

たとえば，図 7 にある最初の規則では，語尾が「н」であり，かつ接続している自立語の語幹末尾が「双母音」であれば，この語尾を助詞「  」に変換する．図 7 にある最後の規則を使うと，日本語の「明後日から」の意味を表す現代モンゴル語「нөгөөдрөөс」が「нөгөөдр」と「өөс」に分割され，さらに語尾の「өөс」が「  」に変換される．

「語幹末尾」の欄が「–」の場合は無条件に語尾を助詞へ変換する．

語幹末尾	語尾	助詞
双母音	нхаа	
	нхоо	
	нхээ	
	нхөө	
н	ийхээ	
	ийхөө	
ж	ийхаа	
	ийхоо	
ш	ийхаа	
	ийхоо	
ч	ийхаа	
	ийхоо	
子音	ийнхээ	
	ийнхөө	
母音	ийнхээ	
	ийнхөө	
н	ыхаа	
	ыхоо	
子音	ынхаа	
	ынхоо	

語幹末尾	語尾	助詞
с,г 以外	даа	
	доо	
	дээ	
	дөө	
子音	оороо	
母音		
子音	өөрөө	
母音		
子音	ээрээ	
母音		
–	тайгаа	
	тэйгээ	
	тойгоо	
–	аасаа	
	ооСоо	
	ээСээ	
	өөСөө	
–	тайг	
	тэйг	
	тойг	

語幹末尾	語尾	助詞
母音	ынхаа	
	ынхоо	
子音	ыгаа	
	ыгоо	
母音	ыгаа	
	ыгоо	
子音	ийгээ	
	ийгөө	
母音	ийгээ	
	ийгөө	
子音	аараа	
母音		
–	тайд	
	тайд	
	тайд	
–	тайгаас	
	тэйгээс	
	тойгоос	
с,г	таа	
	тоо	
	тээ	
	төө	

図 6 現代モンゴル語の語尾を分割して伝統的モンゴル語の助詞に変換する規則（二重格の助詞）

Fig. 6 Rules to separate suffixes of the Modern Mongolian and convert them into postpositional particles of the Traditional Mongolian (double case).

4.2.3 長母音の処理

長母音の処理では，まず，文節の中に長母音が存在するかどうかを検査する．長母音が存在する場合は，長母音の音節対応表を参照して伝統的モンゴル語の表記に変換する．

現代モンゴル語では，長母音を表記する専用の文字がある．しかし，伝統的モンゴル語では，長母音を表記する文字がないため，代わりに 4 種類の表記法がある（2.3.1 項参照）．

そこで，4 種類の表記から適切な表記を選択する必要がある．これには 3 通りの方法がある．まず，変換対象の長母音が語頭の音節にある場合は，**図 8** に示す対応表によって伝統的モンゴル語の音節に変換することができる．

図 8 において，たとえば，長母音「aa」が語頭の音節にあれば，「」に変換される．

次に，語中の位置や前後の文字に関係なく変換される長母音を**図 9** に示す．

語幹末尾	語尾	助詞	語幹末尾	語尾	助詞
双母音	н		子音	оо	
н	ий		母音		
子音	ийн		子音	ээ	
母音			母音		
н	ы		子音	өө	
子音	ын		母音		
母音			–	дээ	
子音	ыг		–	даа	
母音			–	доо	
子音	ийг		–	дөө	
母音			–	таа	
с,г	т		–	тээ	
с,г 以外	д		–	тоо	
子音	аар		–	төө	
母音			–	тай	
子音	оор		–	тэй	
母音			–	той	
子音	өөр		–	аас	
母音			–	ээс	
子音	ээр		–	оос	
母音			–	өөс	
子音	аа				
母音					

図 7 現代モンゴル語の語尾を分割して伝統的モンゴル語の助詞に変換する規則

Fig. 7 Rules to separate suffixes of the Modern Mongolian and convert them into postpositional particles of the Traditional Mongolian.

長母音	音節	長母音	音節
аа		ий	
ээ		өө	
оо		үү	
уу		ёо	

図 8 語頭の音節にある長母音の変換規則

Fig. 8 Conversion rules for the long vowel in the syllable at the beginning of a word.

長母音	音節	長母音	音節	長母音	音節
яа		юу		еө	
еэ		юү		ы	

図 9 無条件に変換される長母音

Fig. 9 Long vowels that can be converted unconditionally.

長母音	第一音節		第二音節以後	
	直前文字	音節	直前文字	音節
аа	ж з		ж з	
	р ш		р ш	
	その他		その他	
ээ	д（л）		д（他）	
	ш		р ш	
	その他		その他	
ий	子音		子音	
уу	子音		子音	
	母音（х）		子音	
оо	т		р ш	
	その他		その他	
өө	子音		子音	
үү	р ш		р ш	
	ж з		ж з	
	その他		その他	

図 10 直前の文字による長母音の変換規則

Fig. 10 Conversion rules for long vowels driven by the previous character.

最後に，図 8 と図 9 のどちらにもない長母音は，その直前にある文字によって伝統的モンゴル語の音節に変換する．この規則を**図 10** に示す．

図 10 では，たとえば，現代モンゴル語の長母音「aa」の直前が「ж」か「з」であれば「」に変換し，「р」か「ш」であれば「」に変換し，それ以外の場合は「」に変換する．

しかし，直前の文字だけで一意に変換することができない場合は，長母音の直前と直後の文字を参照する．図 10 では，変換の条件となる直後の文字を括弧中に示した．すなわち，長母音「ээ」が第 1 音節にある場合は，直前の文字が「д」でかつ直後の文字が「л」であれば「」に変換する．

4.2.4 文字の変換

4.2.2 と 4.2.3 項の処理で変換されていないキリル文字をモンゴル文字に変換する．ここでは，現代モンゴル語と伝統的モンゴル語の対応表（**図 11**）を作成し，利用する．図 11 では，モンゴル文字の字形により混乱を来たさないように独立形や語頭形で表記する．入力テキストの先頭文字から順番に図 11「キリル文字」の欄と照合し，一致した場合は，「モンゴル文字」の欄にある文字に変換していく．本論文では，モンゴル語の習慣表記を尊重してモンゴル文字によって表記し

キリル文字	モンゴル文字	キリル文字	モンゴル文字
а		м	
э		н	
и		р	
о		д	
у		ж	
ө		з	
ү		с	
я		т	
е	(ө) 	х	
	(э,ү)	ц	
ё		ч	
ю	(陽) 	ш	
	(陰)	к	
ы		п	
й		ф	
б		щ	
в		ь	 (я,е,ё)
г			
л		ъ	

図 11 キリル文字からモンゴル文字への変換規則
Fig. 11 Rules to convert Cyrillic characters into Mongolian characters.

ている．しかし，システムの中ではモンゴル文字に対応するローマ字で扱う．

図 11 では，たとえば，キリル文字の「a」はモンゴル文字の「」に変換される．ただし，現代モンゴル語の 1 文字が伝統的モンゴル語で 2 文字に対応する場合がある．たとえば，現代モンゴル語の「я」は伝統的モンゴル語の子音「」と母音「」が結合して「」となる．

また，変換先の文字が一意に決まらない場合がある．その場合は，直前か直後の文字を調べることによって変換後の文字を特定することができる．図 11 は，直前の文字を左側に括弧の中で表記し，直後の文字を右側に括弧の中で表記している．

たとえば，現代モンゴル語の文字「e」は，「親しい」の意味を表す語「эетэй」（eyetei）では「э」の直後なので「」に変換され，その結果「эетэй」は「」（eyetei）に変換される．

現代モンゴル語の音符「ь」は，子音の軟化を示す記号であり，伝統的モンゴル語において弱化母音の代わりになるため，前後の文字によって変換される文字が異なる．たとえば，日本語の「行きましょう」という意味を表す語「явья」では「ь」の直後文字が「я」なので「ь」が「」(o) に変換され，その結果「явья」は「」(yaboya) になる．しかし，日本語の「関係」という意味を表す語「харьцаа」では，「ь」の直後の文字が「я」，「e」，「ё」のどれでもないので「ь」が「」(i) に変換され，その結果は「харьцаа」が「」(haricaga) になる．

伝統的モンゴル文字の字形は，満ら[6] の手法によって自動的に決定する．具体的には，語の先頭に位置する文字は語頭形，末尾に位置する文字は語尾形，他は語中形にする．また，1 つの文字で構成された単語は独立形にする．単語の区切りは空白などの特殊文字で判別する．同じ位置に対して複数の字形をとりうる文字は直前直後の文字によって特定できる．たとえば，子音「」は語中において「」と「」の字形をとりうる．ここで，直後の文字が母音の場合は字形「」をとり，子音の場合は字形「」をとる．

しかし，2.3.1 項の冒頭で説明に使用した母音の「a」には語尾形として「」と「」の 2 種類がある．これらの字形に対しては，満らの手法でも自動的に特定することはできず，彼らが使用したコーパスで使用頻度が高い「」に統一している．

4.2.5 正字法の適用

現代モンゴル語から伝統的モンゴル語へ文字単位で変換するだけでは，伝統的モンゴル語の正字法に違反する音節が生じることがある．

伝統的モンゴル語では弱化母音を表記するのに対して，現代モンゴル語では弱化母音を表記しない．

また，伝統的モンゴル語では語を構成するとき，子音が 3 つ以上連続することはできない．現代モンゴル語の子音には直前か直後に母音を必要とする子音とそうではない子音がある．母音を必要としない子音が，母音を必要とする子音の間に入ると「母音，子音，子音，子音，母音」のように並び，子音が 3 つ連続することがある．

そこで，変換した結果を単語単位で調べ，伝統的モンゴル語の正字法[2)~4)] に従って，欠落した母音を補完する必要がある．ここでは，接続による子音の分類（**図 12**）と補完母音に関する規則（**図 13**）を作成して利用する．

図 12 を使って「後ろに母音だけが接続する子音」の有無を調べて母音が欠落している位置を特定する．伝統的モンゴル語では，単語を構成する母音の発音は母音調和規則に従う．具体的には，ある単語を構成する母音の発音はその単語の第 1 音節と第 2 音節によっ

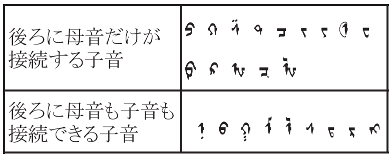


図 12　伝統的モンゴル語における子音の分類
Fig. 12　Classification of consonants in the Traditional Mongolian.

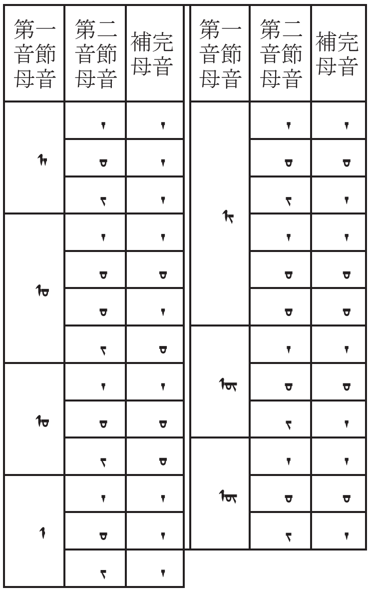


図 13　伝統的モンゴル語の母音補完規則
Fig. 13　Rules to insert vowels in the Traditional Mongolian.

て決まる．

しかし，例外がある．たとえば，形容動詞が未来形であり，かつ語尾が子音「x」の場合は，その後ろに補完する母音は第 1 音節や第 2 音節に関係なく決まる．具体的には，その形容動詞が陽性語なら「ᠣ」であり，陰性語なら「ᠥ」である．ここで，語尾「x」の直前に母音があれば形容動詞の未来形と判断する．

そこで，図 13 の規則を用いて補完する母音を特定する．図 13 において，母音は語中の位置に応じた字形で表記している．そこで，図 3 に示した独立形とは一致しない．

たとえば，図 13 にある最初の規則は，語の第 1 音節の母音が「ᠠ」で，かつ第 2 音節が「ᠠ」であれば，母音が欠落している位置に母音「ᠠ」を補完する．

図 13 では，母音調和規則に従って，第 1 音節と第 2 音節のどちらかが陽性母音か陰性母音の場合は，補完する母音の性もそれと統一している．しかし，第 1 音節と第 2 音節が両方とも中性母音の場合は，補完する母音の性が他の母音と異なることがある．このような場合は，母音調和規則を改めて適応する．具体的には，最初に現れる陽性母音か陰性母音を調べて，補完した母音をこれと同じ性に統一する．

また，上記の規則に従わない例外も存在する．伝統的モンゴル語では同音異義語を弱化母音によって区別する．たとえば，伝統的モンゴル語で発音（hair）を「愛」の意味では「ᠬᠠᠶᠢᠷᠠ」（hairA）と表記し，「砂利」の意味では「ᠬᠠᠶᠢᠷ」（hair）と表記する．しかし，現代モンゴル語では，音声学論を重視し，弱化母音を表記しないため，現代モンゴル語では両方を「хайр」と表記し「愛」なのか「砂利」なのか区別することができない．本来このような語の変換には文脈情報が必要である．しかし，本手法では実験データで使用頻度が高い表記に統一する．文脈に応じた曖昧性の解消は今後の研究課題である．

固有名詞は母音調和規則に従わない場合がある（2.1 節参照）．このような例外には現在対応できていない．

また，正字法に従っても正しく直すことができない例外がある．たとえば，日本語の「人間」を伝統的モンゴル語では「ᠬᠦᠮᠦᠨ」（humun）と表記する．しかし，現代モンゴル語では「хүн」と表記するため，文字変換の結果は「ᠬᠦᠨ」（hun）となる．この語を長母音や母音調和規則に従って修正しても「ᠬᠦᠨ」（hun）のままである．本研究では，このような不規則な語や固有名詞に対処しておらず，今後の研究課題とする．

4.3 伝統的モンゴル語から現代モンゴル語への翻字

4.3.1 単語の抽出

伝統的モンゴル語のテキストは，分かち書きされているため，空白によって機械的に単語を抽出する．

4.3.2 助詞の処理

伝統的モンゴル語の助詞リスト（図 6 と図 7 最右列）を参照して，4.3.1 項で抽出された単語が助詞かどうかを順番に調べる．

現代モンゴル語では助詞が単語と接続しているため，母音調和規則に従って，接続する語幹の性によって助詞の表記が異なる．また，接続する語幹の末尾によっても表記が異なる．

たとえば，助詞「の」は直前にある自立語の末尾が「n」の場合は「ы」と表記し，そうでない場合は「ын」と表記する．また，現代モンゴル語では，第 1 音節の母音が「o」あるいは「θ」の場合は，第 2 音節以後に「o」か「θ」の母音しか入らないという規則がある．よって，活用語尾を接続するときに第 1 音節の母音を調べる必要がある．

また，現代モンゴル語では，活用語尾が語幹に接続

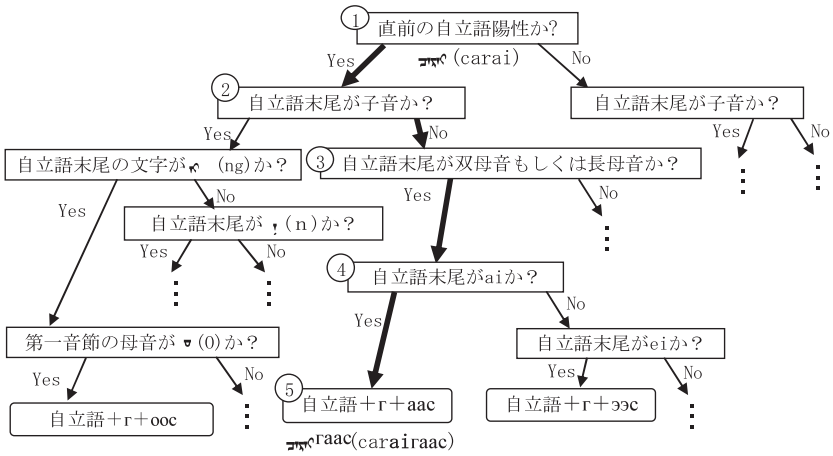


図 14 伝統的モンゴル語の助詞を現代モンゴル語の活用語尾として接続する例：ᠡᠴᠡ（ece）の場合
Fig. 14 Example of connecting a postpositional particles in the Traditional Mongolian as a suffix in the Modern Mongolian: Case of ᠡᠴᠡ (ece).

するとき，語幹の末尾に母音を削除や挿入することがある．そこで，助詞を接続させる語幹（自立語）について，語の性と第 1 音節の母音を調べる．また自立語の末尾がどの文字であるかを調べる．さらに，その自立語の末尾から前の 5 文字を調べて，「削除すべき母音」，「補完すべき母音」，「助詞（語尾）の表記」を特定する．具体的には，「削除すべき母音」と「補完すべき母音」に関する規則を 8 通り，「助詞の表記」に関する規則を 17 通り作成した．これらの規則は，いずれも**図 14** のような決定木で表現することができる．

図 14 は助詞を検出して直前の自立語に活用語尾として接続する規則である．紙面の都合上，残りの 24 通りの規則は割愛する．ここでは，自立語「ᠴᠠᠢ」（carai）に助詞「ᠡᠴᠡ」（ece）を接続する場合を例にとって説明する．

① では，まず助詞「ᠡᠴᠡ」の直前にある自立語が陽性語か陰性語かを調べる．ここで，「ᠴᠠᠢ」（carai）は陽性語なので，次に ② で末尾の文字が母音か子音かを調べる．「ᠴᠠᠢ」（carai）の末尾は母音なので，③ で「ᠴᠠᠢ」（carai）の末尾の母音が双母音かもしくは長母音かを調べる．「ᠴᠠᠢ」（carai）の末尾は双母音なので，④ で「ᠴᠠᠢ」（carai）の末尾が具体的にどの双母音かを調べる．調べた結果は「ᠠᠢ」（ai）なので，⑤ で助詞「ᠡᠴᠡ」を活用語尾「aac」に変換し，「aac」の直前に子音「r」を追加して自立語「ᠴᠠᠢ」（carai）に「raac」を接続する．この時点では自立語は変換されていないので，自立語部分がモンゴル文字のままであり，「ᠴᠠᠢraac」（carairaac）となっている．

4.3.3 長母音の処理

長母音の処理では，まず，語中に長母音を表記する音節があるかどうかを検査する．長母音を表記する音節がある場合は，現代モンゴル語の長母音に変換する．長母音を表記する音節と長母音の対応を**図 15** に示す．

音節	長母音	音節	長母音	音節	長母音
ᠠᠢ	aa	ᠣᠣ	oo	ᠥᠦ	YY
ᠠᠷ	aa	ᠤᠢ	уу	ᠥᠷ	YY
ᠠᠩ	aa	ᠤᠩ	уу	ᠢᠠ	яа
ᠡᠦ	ээ	ᠤᠷ	уу	ᠢᠣ	ёо
ᠡᠷ	ээ	ᠣᠣ	уу	ᠢᠣᠣ	ёо
ᠢᠢ	ий	ᠥᠢ	өө	ᠢᠡ	еэ
ᠢᠷ	ий	ᠥᠷ	өө	ᠢᠡ	еө
ᠤᠷ	оо	ᠥᠷ	өө	ᠢᠤ	юу
ᠤᠦ	оо	ᠥᠦ	YY	ᠢᠥ	юY
ᠤᠩ	оо	ᠥᠩ	YY	ᠢᠪ	ы

図 15 伝統的モンゴル語の音節と現代モンゴル語の長母音の対応
Fig. 15 Correspondence between syllables in the Traditional Mongolian and long vowels in the Modern Mongolian.

図 15 を用いて，「音節」の欄に書かれた伝統的モンゴル語を「長母音」の欄に書かれた現代モンゴル語に変換する．

4.3.4 文字の変換

伝統的モンゴル語と現代モンゴル語の文字数が異なるため子音と母音が 1 対 1 に対応しない．しかし，語中の位置や直前直後の文字によって変換後の文字を特定することができる．

モンゴル文字	キリル文字	モンゴル文字	キリル文字
ᠠ	а	ᠮ	м
ᠡ	э	ᠨ	н
ᠢ	и	ᠷ	р
ᠣ	о	ᠳ	д
ᠤ	у	ᠭ	г
ᠥ	ө	ᠯ	л
ᠦ	ү	ᠰ	с
ᠶ	я	ᠲ	т
ᠧ	е	ᠬ	х
ᡝ	ё	ᠺ	к
ᡠ	ю	ᠪ	п
ᡡ	е	ᠹ	ф
ᡢ	ю	ᠱ	ш

図 16 モンゴル文字からキリル文字へ一意に変換できる文字の対応表
Fig. 16 One-to-one mapping between Mongolian characters and Cyrillic characters.

図 16 はモンゴル文字をキリル文字に変換する規則である．

しかし，モンゴル文字の 1 文字が 2 種類のキリル文字に対応して一意に特定できない場合がある．その場合は**図 17** の規則を使い，語の性，語中の位置や直前直後の文字によってキリル文字に変換する．たとえば，モンゴル文字の「ᠪ」(b) が語頭にある場合は「б」に変換し，語尾にある場合は「в」に変換する．語中にある場合は，その直前の文字が「l」，「m」，「b」，「n」のいずれかであれば「б」に変換し，そうでなければ「в」に変換する．

「キリル文字」の欄が「削除」の場合は該当する母音を削除する．「–」は該当部分の条件は使用しないことを表す．

実際は，人名などの固有名詞は大文字で始まる．しかし，固有名詞を自動的に検出することは難しい．そこで，本手法ではすべて小文字で表記している．

4.3.5 正字法の適用

現代モンゴル語の子音には，単語を構成するときに，直前か直後に必ず母音を必要とする子音「б」，「в」，「г」，「л」，「м」，「н」，「р」と，直前か直後に母音があってもなくてもよい子音（母音を必要としない子音）「д」，「ж」，「з」，「с」，「т」，「х」，「ц」，「ч」，「ш」がある．

ただし，「母音を必要としない子音」は，「母音を必要とする子音」の後ろに接続するときは母音を必要としない．また，子音「с」や「х」の後に子音「т」や「ч」が接続する場合も母音を必要としない．それ以外の場合は 2 番目の子音が直前または直後に母音を必要とする．

モンゴル文字	語の性	語中位置	直前文字	直後文字	キリル文字
ᠪ	–	語頭	–	–	б
		語尾	–	–	в
		語中	l,m, b,n	–	б
			上記以外	–	в
ᠵ	–	語頭	–	i	ж
		語頭以外	–	i以外	з
ᠴ	–	語頭	–	i	ч
		語頭以外	–	i以外	ц
ᠶ	陽	–	j,c,x, n,g	–	ий
			上記以外		ы
	陰	–	–	–	ий
ᠠ	陽	語頭	–	–	я
		語頭以外	母音	–	й
			r, l, m, n		ь
			上記以外の子音		и
	陰	語頭	–	–	и
		語頭以外	j, c, x	–	削除
			上記以外の子音	–	ь
ᠤ	陽	語中	r, l, m, n	y	ъ
			上記以外の子音	–	y

図 17 モンゴル文字からキリル文字への変換規則
Fig. 17 Rules to convert Mongolian characters into Cyrillic characters.

伝統的モンゴル語では，母音を省略しないため，文字単位で変換した結果（4.3.4 項参照）において，母音の欠落はない．現代モンゴル語から伝統的モンゴル語への翻字では母音を補完した（4.2.5 項参照）．それに対して，伝統的モンゴル語から現代モンゴル語への翻字では，母音の欠落はないので補完する必要がない．しかし，余分な母音が入っている場合は，これを削除する必要がある．

よって，伝統的モンゴル語から現代モンゴル語への変換結果を単語単位で先頭から 1 文字ずつ調べ，母音か子音かを調べる．さらに「母音を必要とする子音」か

「母音を必要としない子音」かを調べる．さらに，子音の前後に接続している文字が正字法に従っているかどうかを調べ，母音を必要としない位置にある母音を削除する．また，語の末尾では，子音「г」と「н」の直後に短母音が出現する．「г」と「н」以外の子音は，直後に短母音「и」と音符「ь」以外の母音をとらない．よって，語の末尾に出現している余分な短母音も削除する．

たとえば，伝統的モンゴル語の「ᠮᠠᠨᠳᠤᠲᠤᠭᠠᠢ」（mandotogai）は現代モンゴル語に文字単位で変換すると「мандутугаий」（mandotogai）になる．この変換結果を正字法で修正する過程を以下に示す．

先頭の子音「м」（m）は母音を必要とするので次の母音「a」（a）は必要である．子音「н」（n）は母音を必要とする．しかし，直前に母音「a」（a）があるので直後に母音を必要としない．子音「д」（d）は母音を必要としない子音で，かつ直前に母音を必要とする子音「н」（n）があるので直後に母音を必要としない．よって，子音「д」（d）の直後にある母音「y」（o）を削除する．子音「т」（t）は母音を必要としない．しかし，子音「д」（d）の直後にあるので母音を必要とする（「т」の直前にある「y」はすでに削除されている点に注意）．よって，子音「т」（t）の直後にある「y」（o）は必要である．子音「г」（g）は母音を必要とする子音であり，直前に母音があるので直後に母音は必要としない．しかし，双母音もしくは長母音の場合は必要なので，最後の双母音「ай」（ai）も必要である．最終的に，「мандутугаий」（mandotogai）は「мандтугаий」（mandtogai）に修正される．

現代モンゴル語から伝統的モンゴル語の翻字における正字法（4.2.5 項参照）では，母音調和規則を適用して補完された母音の性を統一した．しかし，ここでは母音を削除するだけなので，母音調和規則を改めて適用する必要はない．

5. 翻字手法の評価実験

5.1 実験の方法と結果

現代モンゴル語から伝統的モンゴル語への翻字では，新聞記事の Web サイト☆から入手した 10 記事を実験データとして使用した．

伝統的モンゴル語から現代モンゴル語への翻字では，満ら[6] が開発した伝統的モンゴル語入出力インタフェースを用いて人手で入力し，電子化した新聞記事（内蒙古日報）31 件を使用した．実験結果を**表 1** に示す．正解判定は，著者以外のモンゴル人 1 名によって行った．

☆ http://onigoo.olloo.mn

表 1 翻字の実験結果
Table 1 Experimental results for transliteration.

	現代→伝統		伝統→現代	
記事数	10		31	
語数	延べ	異なり	延べ	異なり
	6,943	2,223	14,435	2,895
正解した語数	5,596	1,827	12,343	2,233
精度（%）	80.6	82.2	85.5	77.1

表 1 では，テキスト中の語数を「延べ」と「異なり」で別々に数えた．「延べ」はテキスト全体での翻字精度である．「異なり」は高頻出語と低頻出語を同等に評価する場合の翻字精度である．

現代モンゴル語から伝統的モンゴル語への翻字精度は述べ語で 80.6%，異なり語で 82.2%であった．伝統的モンゴル語から現代モンゴル語への翻字精度は述べ語で 85.5%，異なり語で 77.1%であった．さらに，変換結果のテキストに誤りがあっても内容の理解には支障のないことが分かった．すなわち，本手法は実用的なレベルに到達している．

5.2 誤り分析

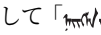
翻字結果のテキストは内容理解に支障がなかった．しかし，現代モンゴル語から伝統的モンゴル語への翻字では，延べ語の 19.4%，異なり語の 17.8%は誤りであった．誤った 396 語（異なり）を分析した結果，以下に示す（a）〜（c）の原因によることが分かった．

（a） 略語 24.7%（98/396）
（b） 固有名詞 48.5%（192/396）
（c） 語源の違い 26.8%（106/396）

（a）は伝統的モンゴル語では省略表記をしないため，現代モンゴル語の略語を文字単位で変換すると意味不明の文字列になることに起因する．たとえば，日本語の「国会」を現代モンゴル語では「Улсын Их Хурал」と表記し，「УИХ」のように略記する．しかし，伝統的モンゴル語では略記しないので，「УИХ」を変換した結果は意味不明の文字列になる．

（b）は，固有名詞が母音調和規則に従わないことに起因する．本手法では，1 つの単語に含まれるすべての母音が母音調和規則に従って，中性母音以外は陽性か陰性のどちらかに統一される．たとえば，「バトムンク」という人名を現代モンゴル語で「Батмөнх」（batmvngh）と表記し，伝統的モンゴル語で「ᠪᠠᠲᠤᠮᠥᠩᠬᠡ」（batomvnghe）のように表記する．しかし，変換結果に伝統的モンゴル語の正字法を適応すると「ᠪᠠᠲᠤᠮᠤᠩᠬᠠ」（batomongha）のように第 1 音節の母音の性に合わせてすべての母音が陽性に統一されてしまう．しかし，母音調和規則に従う固有名詞もある．たとえば，

「амаржээ」(amarjee) という現代モンゴル語の人名は，母音調和規則に従わない．それに対して，伝統的モンゴル語では母音調和規則に従って，「  」(amarjai) と表記するため，母音調和規則を適用しないと誤って翻字されてしまう．そこで当問題は，「固有名詞には母音調和規則を適用しない」というような単純な方法では解決しない．

(c) は，言葉を借用する原言語の違いや音訳と意味訳の違いに起因する[7]．たとえば，「社会主義」という言葉を，現代モンゴル語ではロシア語を音訳して「социализм」(socialism) と表記し，伝統的モンゴル語では中国語を音訳ではなく意味訳して「  」(neigem jirum) と表記する．両単語の発音は異なるため，翻字すると目的語で意味不明の文字列になってしまう．

伝統的モンゴル語から現代モンゴル語への翻字で誤った 662 語（異なり）を分析した結果，以下に示す (d)〜(f) の原因によることが分かった．

(d) 長母音の検出誤り 45.0%（298/662）
(e) 短母音に関する慣習の違い 31.4%（208/662）
(f) 双母音に関する慣習の違い 23.6%（156/662）

(d) は，伝統的モンゴル語の長母音を表記する音節が長母音以外の表記にも使われることに起因する．たとえば，語「  」(egedeng) では音節「ge」が短母音「e」を長音化して，長母音「ээ」を表記する．しかし，別の語では音節「ge」が短母音「e」の直後に接続しても短母音「e」と独立して発音する場合がある．たとえば，語「  」(egexig) では「ge」が短母音「e」を長音化せずに，つづり読みする．

(e) は，伝統的モンゴル語では，慣習上，長母音を短母音で表記する場合があることに起因する．たとえば，日本語の「空気」の発音は「agaar」であり，伝統的モンゴル語では「  」(agar) と表記する．しかし，現代モンゴル語では「агаар」(agaar) と表記する．

(f) は，伝統的モンゴル語では，慣習上，長母音を双母音で表記する場合があることに起因する．たとえば，日本語の「兎」を伝統的モンゴル語では「  」(taolai) と表記する．しかし，現代モンゴル語では「туулай」(toolai) と表記する．

以上，(a)〜(f) の誤りに根本的に対処することは難しい．現状では，例外的な単語として辞書に登録し，翻字処理の際に参照するよりほかに方法がない．

6. おわりに

本論文は，伝統的モンゴル語と現代モンゴル語のテキストを規則によって相互に変換する翻字手法を提案した．評価実験の結果，現代モンゴル語から伝統モンゴル語への翻字精度は 80.6%，伝統的モンゴル語から現代モンゴル語への翻字精度は 85.5%に達した．

本手法によって，モンゴル国と内モンゴルの情報交換を活発化することが期待できる．

また，電子化テキストデータが乏しい伝統的モンゴル語のテキスト処理に関する研究を促進する効果も期待できる．

参 考 文 献

1) ガラサンポンサグ：モンゴル国のキリル文字正字法，内蒙古人民出版社，呼和浩特市 (2001).（伝統的モンゴル語）
2) 清格尔泰：現代モンゴル語文法，内蒙古人民出版社，呼和浩特市 (1999).（伝統的モンゴル語）
3) トゴ：モンゴル語文法概要，内蒙古少年児童出版社，呼和浩特市 (1986).（伝統的モンゴル語）
4) 那任巴図：現代蒙古語，内蒙古大学出版社，呼和浩特市 (1995).（伝統的モンゴル語）
5) 中里致元，生出恭治：現代モンゴル語の異種表記法の相互変換システムの構築に向けて，情報処理学会研究報告，2002-CH-53, pp.41–46 (2003).
6) 満 都拉，藤井 敦，石川徹也：伝統的モンゴル語の電子化方式とテキスト検索への応用，電子情報通信学会論文誌 D-II，Vol.J88-D-II, No.10, pp.2102–2111 (2005).
7) 満 都拉：モンゴル語専門用語の由来，第 13 回（2000 年度）専門用語研究シンポジウム，pp.13–20 (2000).

(平成 17 年 11 月 8 日受付)
(平成 18 年 5 月 9 日採録)

満 都拉

1986 年内蒙古農牧学院卒業，同年同大学附属図書館に就職．2002 年図書館情報大学大学院情報メディア研究科博士前期課程修了．同年同大学院博士後期課程に進学．現在，大学統合にともない筑波大学大学院図書館情報メディア研究科に在籍中．

藤井　敦（正会員）

1993 年 3 月東京工業大学工学部情報工学科卒業．1998 年 3 月同大学大学院博士課程修了．図書館情報大学助手を経て，現在，筑波大学大学院図書館情報メディア研究科助教授，博士（工学）．自然言語処理，情報検索，音声言語処理，Web マイニングの研究に従事．電子情報通信学会，人工知能学会，言語処理学会，Association for Computational Linguistics 各会員．

石川　徹也（正会員）

1971 年 3 月慶應義塾大学大学院修士課程（図書館情報学専攻）修了．富士フイルム（株）足柄研究所，図書館短期大学，図書館情報大学，筑波大学を経て，2006 年 4 月から東京大学史料編纂所前近代日本史情報国際センター特任教授．歴史情報学の研究に従事．筑波大学名誉教授．筑波大学大学院図書館情報メディア研究科客員教授．工学博士．ACM，言語処理学会等会員．