【9】

氏　名（国籍）	満　都　拉（中　国）（まん　と　ら）
学位の種類	博　士（情報学）
学位記番号	博　甲　第　4823　号
学位授与年月日	平成20年6月30日
学位授与の要件	学位規則第4条第1項該当
審査研究科	図書館情報メディア研究科
学位論文題目	**伝統的モンゴル語と現代モンゴル語の電子化と双方向翻字手法の研究**

主　査	筑波大学教授	石　塚　英　弘
副　査	筑波大学教授	田　中　和　世
副　査	筑波大学教授	杉　本　重　雄
副　査	東京大学教授	石　川　徹　也
副　査	東京大学教授	堀　　浩　一

論文の内容の要旨

本論文は，異なる国に居住し，異なる文字を使用しているモンゴル族が，他国で公開されたモンゴル語テキストを自分が読める文字のテキストに変換して読めるようにすることを目的として，これまで困難とされてきた伝統的モンゴル語テキストの電子化方式を確立し，他のモンゴル語との間の双方向翻字手法を開発した研究の論文である。

第1章では，研究の背景と目的について次のように述べている。現在，モンゴル語を使用するモンゴル族はモンゴル国に約270万人，中華人民共和国では内蒙古自治区に約421万人と新疆ウィグル自治区に約17万人の合わせて約438万人と，2つの国に別れて居住している。またモンゴル語とそのテキスト表記文字は，モンゴル国では「現代モンゴル語」と「キリル文字 」（ロシア語キリル文字に特別に2文字が追加されているため，正式にはモンゴル語キリル文字）が使用されている。また，中華人民共和国では内蒙古自治区で「伝統的モンゴル語」と「伝統的モンゴル文字」が，新疆ウィグル自治区で「トドモンゴル語」と「トド文字」が使用されている。その結果，モンゴル族の人間であっても，その人が知らない文字で書いてあるモンゴル語のテキストは読めないという事態が生じている。インターネット上に公開されていても，たとえば，伝統的モンゴル文字で書かれた伝統的モンゴル語テキストはモンゴル国のモンゴル族には読めない。伝統的モンゴル文字を知らないためである。また，伝統的モンゴル文字を使用する伝統的モンゴル語テキストはコンピュータ処理が難しいため，電子化が遅れている。この現状を改善するため，具体的にはインターネット上の他国のモンゴル語テキストを自国で使用している文字で読むことができるようにすることを本研究の目的として，本論文の著者（中華人民共和国内蒙古自治区出身）は，伝統的モンゴル語の電子化方式を確立し，伝統的モンゴル語と現代モンゴル語の双方向翻字手法を開発すると述べている。

第2章では伝統的モンゴル語と現代モンゴル語の言語学的特徴をそれぞれ文献に基づいて紹介し，両モンゴル語の共通点と相違点を述べている。具体的には，文字の構造，語の表記規則，発音と表記の関係などであり，これらは後に述べるテキストの電子化手法，双方向翻字手法が満たすべき機能要件に直接関係する事実である。たとえば，伝統的モンゴル文字は表音文字ではあるが，1文字だけを書く時は「独立形」が用い

られ，その文字が語の頭に在る時は「語頭形」語の中に在る時は「語中形」語尾に在る時は「語尾形」と異なる形の字（字形）が使用され，また語の中で母音調和規則と正字法（単語を構成する文字同士の連結に関する規則）に従って字形が決まる場合があること，同形異音文字が存在することなどの特徴を持つ独特な文字である。一方，現代モンゴル語ではキリル文字を使用するため，語頭形，語中形などは消失している。また，伝統的モンゴル語と現代モンゴル語の相違点として，助詞の表記（伝統的モンゴル語では自立語と助詞を分ち書きするが，現代モンゴル語では助詞は接尾辞として自立語に結合する。），長母音の表記，弱化母音の表記，正字法を挙げている。

第 3 章「伝統的モンゴル語の電子化」では，まず既存の電子化手法「文字コードによる電子化」と「ローマ字転写による電子化」を概観し，前者には同形異音文字を区別できない欠点があり，後者には 2 つの手法があるが，一方には同形異音文字を区別できない欠点が，他方には曖昧性が生じる欠点や英数字の混在するデータを扱えない欠点があることを指摘している。これに対して，本研究の手法はローマ字転写方式を改良した手法であり，一文字を 1 つの発音に対応させ，かつ，転写の際に数字を用いない手法であるため，前述の欠点を全て解決し，2 章で述べた要件を満たすと述べている。

本手法による電子化データは Unicode によるデータとは双方向の完全互換性があること，また既存の電子化手法データへの変換は可能で，その逆変換は不完全になること，その原因は既存の電子化手法データに一部情報の欠落が在るためと指摘している。

また，伝統的モンゴル語の入出力インターフェースも開発し，テキストの入力，保存，出力を可能にした。そして，これを用いて伝統的モンゴル語で書かれた新聞の 31 記事を電子化する検証実験を行い，31 記事に存在する異なり文節 4237 の全てが曖昧性無く電子化されたと述べている。

第 4 章「伝統的モンゴル語と現代モンゴル語の双方向翻字手法」では，まず既存の研究では伝統的モンゴル語の単語を現代モンゴル語の単語に翻字する方法を検討した論文があるが，手法の提案までは至っていないと述べ，次いで本研究では，原言語を伝統的モンゴル語，目的言語を現代モンゴル語とする翻字手法と，原言語を現代モンゴル語，目的言語を伝統的モンゴル語とする翻字手法の 2 つを開発したと述べている。翻字手法の手順は，原言語テキストを源として，1）単語の抽出，2）助詞の処理，3）長母音の処理，4）文字の変換，5）目的言語の正字法の適用，で目的言語テキストに変換する。個々の処理内容は原言語が伝統的モンゴル語か現代モンゴル語かによって異なるため，それぞれ詳しく述べている。たとえば，原言語が伝統的モンゴル語の場合の助詞の処理では，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尾辞として結合する等々，具体的に述べている。

また，開発した双方向翻字手法をプログラム化し，伝統的モンゴル語新聞記事 31 件（異なり文節数 4237，延べ文節数 14606）と現代モンゴル語新聞記事 50 件（異なり文節数 6809，延べ文節数 24100）を用いて検証している。伝統的モンゴル語から現代モンゴル語への翻字の場合，異なり文節では 69.2%，文節総体では 83.2% が正しく翻字され，残りは正字法に従っていないものの大部分が読めば理解可能なテキストになっていると述べている。現代モンゴル語から伝統的モンゴル語への翻字の場合，異なり文節では 62.2%，文節総体では 80.6% が正しく翻字され，残りは正字法に従っていないものの大部分が読めば理解可能なテキストになっていると述べている。正解・不正解は個々の文法規則に基づいて判定されており，翻字の各処理ごとに正解・不正解の数とその分析が述べられている。また，現代モンゴル語から伝統的モンゴル語への翻字において同音異形語，同音異義語を区別できない場合があることについて考察し，その原因は伝統的モンゴル語では区別する情報を字形として保持しているが，現代モンゴル語ではそれを無くしたためであると述べている。なお，文法規則ごとの正解･不正解の実例は付録に 100 ページを越える表として示されている。

第 5 章は本論文の結論であり，本研究は当初の目的を達成した旨が書かれている。

審 査 の 結 果 の 要 旨

本論文の著者は，インターネット上の他国のモンゴル語テキストをモンゴル族が自国で使用している文字で読むことができるようにすることを研究目的として，伝統的モンゴル語の電子化方式を確立し，伝統的モンゴル語と現代モンゴル語の双方向翻字手法を開発すると述べている。自国のテキスト表記にどの文字を使用するかは，その文字の教育方針と併せて，その国の基本政策であり，国民の文化でもあるため，インターネットが盛んな現在，本研究の目的はよく理解できる。また，目的を達成するために行った 2 つの研究についても理解できる。伝統的モンゴル語は伝統的モンゴル文字と文法規則が複雑なため電子化が遅れていること，双方向翻字によるテキスト変換は両国の基本政策と文化に適合するため，両国に受入れ易い現実的な方策であることがその理由である。

第 3 章に書いてある「伝統的モンゴル語の電子化」では，複雑な構造を持つ伝統的モンゴル語文字と伝統的モンゴル語が有する難しい問題点を分析した結果，新しい手法を考案し，この手法が既存の電子化手法の欠点を解決できることを論理的に示すとともに検証実験によって本手法の有用性を明らかにしたことは評価できる。また，本手法で電子化したデータは Unicode によるデータと双方向の完全互換性があることも有用である。さらに，本手法を電子情報通信学会論文誌に発表した時期が 2005 年であるのに対して，Unicode による伝統的モンゴル語テキストの表示・編集等の実用化が Windows 上では 2007 年の Windows Vista の発売まで実現しなかった事実は本手法の先進性を示しており，評価に価する。

第 4 章に述べられている「伝統的モンゴル語と現代モンゴル語の双方向翻字手法」は複雑なモンゴル語文法の中の翻字に関する部分を自然言語処理の観点から抽出して双方向翻字手法を開発した点に新規性があり，高く評価された。また，本翻字手法により生成したテキストには一部に正字法に反する箇所が在るものの，その部分も含めて人が読めば理解できるテキストであることは，第 1 章で述べてある研究の目的を達成したことを示している。また，モンゴル語文法に基づいて正解・不正解を厳密に判定し，本手法の有用性と限界を明示した点も評価できる。

なお，本手法と実用システムの位置付けに関して審査委員の間で議論があった。検討の結果，1）モンゴル語特有の正字法に起因する問題を除けば精度の高い翻字手法であるが，正字法が存在する以上，実用システムとしては正字法の問題を無視できないため，このままでは翻字システムとまでは言い難いこと，2）実用システム化には，正字法適用処理における翻字精度向上技術，また現代モンゴル語から伝統的モンゴル語への翻字システムの場合には同音異形語，同音異義語を区別するための機能，例えばオペレータとの間の対話選択機能，が必要と考えられるが，それにも拘らず本手法は基盤となる手法として高く評価できることの 2 点で全員が一致した。また，双方向翻字の精度が上がれば，伝統的モンゴル語の電子化と組み合わせて，両モンゴル語の横断的情報検索も将来可能になるとの意見が複数あった。

以上，本研究は当初の目的を達成するとともに，技術的に新規性があり，有用性が高い。また本研究の成果は，伝統的モンゴル語テキストに関する様々な情報処理の展開，伝統的モンゴル語テキストと現代モンゴル語テキストのインターネット上の横断的情報検索システムへの発展など，モンゴル語テキストの国を越えた情報共有･共用の実現に大いに寄与しうるものであり，この分野の研究開発に道を拓くものと期待できる。従って学位論文審査委員全員により論文審査合格と判定された。

よって，著者は博士（情報学）の学位を受けるに十分な資格を有するものと認める。