

中国内蒙古における蒙古語の電子計算機 4V-1 処理の研究の現状

額爾敦¹ 瀧口伸雄² 小谷善行² 西村恕彦²
¹中国吉林廣播電視大学 ²東京農工大学工学部

1. はじめに

新時代の計算機技術の一つは、マンマシン・インタフェースとして自然言語や自然言語に近い言語を採用することである。蒙古民族にとってはそれは蒙古語の採用である。民族の言語の立場からみると蒙古語情報処理技術の確立が必要であり、また、世界で利用可能になるための枠組の提供ということが必要である。この両方面をいかに統一するか、計算機アーキテクチャや入出力装置、システムソフトウェアやアプリケーションソフトウェアなどがどうあるべきか、著者はいろいろな疑問と興味を持って、中国での蒙古語研究の中心である内蒙古自治区の首都フフホト市を1993年3月末ごろ訪問し、調査した。本稿では、中国内蒙古における蒙古語の電子計算機処理の研究の現状を紹介する。

2. 内蒙古における研究機関の概要

内蒙古自治区では、主として、次の研究機関で蒙古語処理の研究が行われている。

- ① 内蒙古大学蒙古語研究所
- ② 内蒙古大学電子研究所
- ③ 内蒙古科学技術計算センター
- ④ 内蒙古大学新技術開発公司

中国文字情報処理研究会^[1]が1981年6月天津に成立した。本研究会には基礎理論、漢字情報処理システム、漢字コード編集、漢字情報処理用装置、自然言語処理などの研究専門委員会が組織されている。

1982年から全国各省における少数民族言語文字研究者が各々の民族言語、文字情報処理に関する研究を開始した。内蒙古科学技術計算センター、内蒙古大学蒙古語研究所はほとんど同じ1983年から研究を始めた。内蒙古科学技術計算センターはシステムの構成や入出力装置に関する研究、内蒙古大学蒙古語研究所は蒙古語の言語学的研究の計算機応用に関する研究、内蒙古大学電子研究所は蒙古語の入出力方式に関する研究、内蒙古大学新技術開発公司是蒙古文字印刷用フォントおよび印刷システムの研究などを行っている。

3. 蒙古語の電子計算機処理の研究

3.1 翻字システムの開発

蒙古語における翻字は、フドム文字（図1）とハルハ文字（図2）間双方向の翻字システムの開発が行われている。

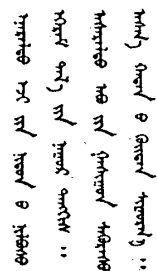


図1 フドム文字

ЭНЭРЭЛТ ЭЭЖИЙН ННДНИЙ НУЛИМС
 ЭЭРЭМ ТАЛИЙН НУУР ТАНГИС.
 АСРАЛТ ААВИЙН ГАШУУН СҮҮРС
 АСГА ХАДНИЙ ХҮЙТЭН ШУУРГА.

図2 ハルハ文字

ハルハ文字をフドム文字に翻字するシステムをSHBPという^[2]。このSHBPの開発は内蒙古大学蒙古語研究所が1986年から行い、1987年の9月に最初のバージョンが完成し、運用され始めた。SHBPはAT&T製3B2/310型マシン上のUNIXシステムでC言語を用いて開発され、現在IBMパソコンに移植されている。

SHBPは、原稿入力、辞書参照、単語の書換え、結果

Current State of the Art of Mongol Computer

Processing in Inner Mongolia of China: Eerdun, Nobuo

Takiguchi, Yoshiyuki Kotani, Hirohiko Nisimura,

Tokyo University of Agriculture and Technology

2-24-16 Naka-cho, Koganei, Tokyo, 184, Japan

出力の4つの部分から構成されている。

① 原稿の入力

原稿の入力としては、対話型入力方式とバッチ型入力方式との2種類がある。対話型入力方式とは、原稿を入力した後すぐに翻字操作を行い、直接翻字結果を得る方式である。バッチ型入力方式とは、原稿を入力した後、原稿に対して適当な編集手続きをして、その編集された二次原稿を翻字する方式である。この2種類の方式では、いずれにしても単語の語幹と語尾を専用符で区切り入力する。

② 辞書参照

SHBPの辞書は、ハルハ文字とフドム文字それぞれの単語の語幹と語尾の集合から構成される。SHBPの中には、ハルハ単語の辞書、フドム単語の辞書（基本辞書）、ハルハ文字の同形異義語の辞書、ハルハ文字の同形異義語に対応するフドム文字の辞書（動詞辞書、名詞辞書）、ハルハ文字単語の語尾辞書、ハルハ文字に使われている外来語辞書、ハルハ文字専用語辞書、ハルハ文字の外来語と専用語に対応するフドム文字単語の辞書などの辞書が含まれている。

辞書引き段階では、語尾変化規則の定義、数詞の定義、語尾表、外来語の翻字形式などの初期化、句読点の定義、語幹の判断、語尾の判断などの作業を行う。

③ 出力結果

フドム文字に翻字された結果をスクリーン上に表示する、または印刷して出力する。

④ 結果についての評価

フドム文字に翻字された結果についての評価は、翻字された単語または文の正確性、システムの信頼性および移植性などの面から評価する。

⑤ システムの利用

このシステムで原稿の入力は、ラテン文字により行い、翻字する操作は、メニューのコマンドを用いて行う。

逆方向にフドム文字をハルハ文字に翻字するシステムは、1992年から内蒙古自治区と蒙古人民共和国との協同で開発中である。

3.2 蒙古語のデータベース

内蒙古大学蒙古語研究所は言語学的研究を目的として、

1986年から蒙古語のデータベースを作り始め、1990年12月までに100万単語のデータベースを作成した。このデータベースは、中国国家社会科学基金の援助を受けて、500万単語まで拡張される計画である。

このデータベースは4つに分けられ、蒙古語文教科書類が50.3%、政治文章類が20.9%、文学芸術類が19%、新聞文章類が9.8%を占めている。

3.3 多種類文字の入出力システム

このシステムは内蒙古大学蒙古語研究所の電子計算機室が開発し、1987年に完成した。このシステムを利用して満州、喫丹、ロシア、英、ブリヤート、カルムイク、ハルハ、パスパ、回鶻、トド、フドムなどの各種類の文字の入出力ができる。

3.4 蒙古語の印刷システム

蒙古語の印刷システムを内蒙古大学新技術開発公司与北京大学新技術開発会社が協同開発し、それを蒙古語の「華光(HUA GUANG)」システムという。

このシステムを利用してフドム、トド、ハルハ、ブリヤート、キリルなど5種類の文字の入力および印刷ができる。1990年から内蒙古自治区、シンジエンウィグル自治区は新聞や雑誌の印刷に使用し始めている。

3.5 その他のプロジェクト

内蒙古大学蒙古語研究所では、蒙古単語の品詞推定システムの開発を中国国家自然科学基金に、そして、蒙古語の単語の綴りの入力システムの開発を内蒙古自治区自然科学基金にそれぞれ援助申請している。

4. おわりに

蒙古文字は右横書きのアラム文字に起源するが、長い歴史の中で独自の発展を遂げて来た。電子計算機技術によって蒙古言語の処理を行うことが、蒙古族の固有の言語、文化の発展およびこれからの世界各民族の相互文化交流にも大きな社会的影響をもたらすに違いない。

参考文献

- [1] 趙珀璋・徐力, 中文計算機処理, 宇航出版社, 1987
- [2] 華沙宝, ハルハ文字をフドム文字に翻字するソフトウェア SHBP, 内蒙古大学科学研究雑誌, 1988