

マイナー言語の入力方針案 モンゴル文字資料を一例に

平成16年10月28日

平成16年度第1回総合目録データベース実務研修

京都大学附属図書館情報管理課システム管理掛 長谷川 裕子

新潟大学附属図書館情報管理課図書情報係 永井 登志江

大阪外国語大学附属図書館利用部門 澤本 亜希

大阪大学サイバーメディアセンター電子図書館掛 久保山 健

1. はじめに

NACSIS - 新 CAT/ILL が平成12年に多言語対応したことにより、中国語簡体字、ハングル、アラビア文字での検索・登録が可能になった。これらの資料は所蔵館の間で NACSIS-CAT 登録の必要性が高まり、NII により規則等が制定・修正され、現在の運用が可能になったものである。

一方で、多言語未対応のマイナー言語資料を所蔵している参加組織が存在し、既存の規則等では判断がつかず、NACSIS - CAT に登録できない、もしくは登録されている書誌レコードの妥当性について作成館と調整できない、といったケースがあるのも事実である。

当グループの報告においては、上記のような背景をもとに、多言語未対応の資料の入力方針について考察する。ただし、ここでは多言語対応していないマイナー言語資料の一例としてモンゴル文字資料を取り上げる。

なお、モンゴル文字資料を取り上げるのは、当グループ員の所属する参加組織の中に、モンゴル文字資料を NACSIS - CAT に登録する際に問題が発生しているという現状を踏まえたことによる。

2. モンゴル文字資料の現状

2.1 モンゴル文字とモンゴル文字資料

モンゴル語とは、モンゴル民族が母語とする言語であり、ロシアのキリル文字を使用した文字と、モンゴル文字（蒙古文字）の主に2種類存在する。ここでは、モンゴル文字と呼ばれる縦書きで書かれた資料について取り上げることとする。

モンゴル文字は表音文字である。表記は縦書きで、行の並びは左から右である。一つの言葉は1本の縦線でつながったように書かれ、言葉ごとに分かち書きがなされる。それぞれの文字は、言葉の中の使う位置によって語頭形、語中形、語末形の3種類に変化する。

モンゴル文字資料は主に、中華人民共和国内蒙古自治区で発行されている。グループ員

がこれまで扱った例によると、内蒙古自治区で発行された資料の情報源には、表紙・標題紙の表示はモンゴル文字、奥付の表示は中国語簡体字、という例が多く見られる。

<参考>

モンゴル語：キリル式文字とモンゴル文字の対応表（一部抜粋）

大文字	小文字	字体名	蒙古文字 (語頭形)	転写(例)
		ah	ᠠ	a
		beh	ᠪ	b
		weh	ᠪ, ᠠ	b/w
		geh	ᠭ, ᠭ	G, / g
		deh	ᠳ, ᠳ	d
		yeh	ᠶ, ᠶ	ye, yo
		yaw	ᠶ	yo

(「言の葉」>「キリル式文字」<http://mariyot.ld.infoseek.co.jp/cyrill.htm>)

2.2 NACSIS-CATにおける入力書誌状況

平成16年10月の当グループの調査時点において、NACSIS-CATにタイトル言語がモンゴル語として登録されている書誌は1,548件である。これが多いか少ないかは一概に言えないだろうが、平成13年に行ったNIIの「外国語図書の所蔵状況等に関する調査結果」(NACSIS-CAT/ILL ニュースレター5号6p 2001.12.20)によると、当時の調査に回答した57機関が所蔵するモンゴル語資料は11,297冊、年間増加冊数は374冊であった。

登録されているモンゴル語資料を見ると、TRフィールドの入力例は下記の3通りであり、登録数など詳細も併せて示す。

<TRフィールド入力例>

- ・キリル文字 (BA21109602)
? ????? ????? ????? / ??? . ??????????
- ・ローマ字翻字 (BA59717597)
Küke sudar : terigün debter / Injenasi
- ・中国語 (BA38030744)
蒙古? 科尔沁土? 研究 / 查干哈[達]著

<TRフィールド入力方式別内訳>

- ・キリル文字 1,222件

- ・ ローマ字翻字 166 件
- ・ 中国語 160 件

上記の例のうち、キリル文字以外で入力された「ローマ字翻字」、「中国語」の合わせて 326 件が、モンゴル文字資料の書誌であり、両者の比率はほぼ同じであった。表紙・標題紙の表示はモンゴル文字、奥付の表示は中国簡体字、という資料のあることを考慮すると、中国語で入力されているレコードに関しては標題紙のモンゴル文字を採用せず、奥付記載の中国語をタイトルとして記述しているものが相当数になるであろう。

2.3 参考：早稲田大学では

参考として早稲田大学提供の「早稲田大学図書館蔵モンゴル文文献目録データベース」におけるモンゴル文字文献書誌情報は、「タイトル」、「著者・编者」、「出版社」、「出版地」がローマ字翻字形と中国語の 2 通りの方法で記載されており、双方からの検索が可能である。ローマ字翻字形にする際は、書名の最初の語や固有名詞を大文字で表すことはせず、全て小文字での表記となっている。また、データベースには資料の表紙画像が収録されており、検索後の詳細表示画面で実際のモンゴル文字を確認することもできる。

< 早大：モンゴル文献検索結果例 >

- ・ タイトル：nomci qatun-u geser-ün tu?uji.
- ・ タイトル（漢字）：? 木其哈敦格斯?

参考 URL：

早稲田大学図書館蔵モンゴル文文献目録データベース

<http://www.littera.waseda.ac.jp/mongol/>

3. これまでの多言語入力規則

3.1 制定経緯

目録業務システムで新 CAT/ILL が稼動したのは平成 9 年であり、その後平成 12 年より目録システムの多言語対応が始まった。UCS が一般のパーソナルコンピュータへも浸透し、多言語を入力できる環境が整ってきた点が大きい。

多言語資料の規則制定としてまず俎上に上ったのは中国語資料である。歴史的背景を鑑みても資料の蔵書数が最も多く、暫定を含め既存レコードも多い中国語が選ばれたものと推定される。

(1) 中国語資料

平成 10 年 10 月に「中国語資料の取扱い(案)」、「中国語資料用コーディングマニュアル(案)」が制定された。このうち、取扱い(案)は平成 11 年 12 月発行の「目録情報の基準 第 4 版」に吸収され、廃止された。

制定前に、参加館へのアンケート実施や入力事例を分析した等の準備は規則等からは伺えないため不明である。しかし、先行入力館として東京大学、京都大学等の中国図書入力マニュアルを参考にしたと推測される。

(2) 韓国・朝鮮語資料

韓国・朝鮮語資料については、平成 14 年 1 月に「韓国・朝鮮語資料の取扱い」「同解説」が制定された。検討段階で NII は平成 12 年 11 月に 1,211 の参加組織へ韓国・朝鮮語資料の所蔵状況、組織化状況についてアンケートを行った(回答 1,021 組織、回答率 84%)。特に翻字法についての結果分析は、「取扱い」へ大きな影響を与えている。

(3) アラビア語資料

「アラビア語資料に関する取扱い及び解説」、「コーディングマニュアル(アラビア文字資料に関する抜粋集)」が NII の図書館情報委員会で承認されたのは平成 14 年度のことである。NII では特に参加館へのアンケート実施は行わなかったようであるが、議事録によると平成 9 年に東京大学にて採択された科学研究費補助金の「イスラム地域研究」におけるアラビア文字文献入力規則ワーキンググループの提言がその発端となったことが分かった。このプロジェクトでは国内の主要所蔵機関へアンケートを行い、意見を調整し平成 14 年 3 月に NII へ提言を行った。NII では平成 15 年 7 月発行の NACSIS-CAT/ILL ニュースレターで、本格運用を通知している。

3.2 各言語の要素比較

各言語のコーディングマニュアルや取扱い等から、記述に関する要素を比較する。

それにより、モンゴル文字資料の入力方針案の検討材料とする。

なお、ここでは日本目録規則 1987 年版を NCR87、英米目録規則第 2 版を AACR2 と呼ぶこととする。

(1) 目録規則等

	適用範囲	適用する目録規則	目録用言語
中国語	(特に規定なし)	NCR87 (*1)	日本語 * 中国語の選択も可能 (「CM案」)
韓国・朝鮮語	(特に規定なし)	NCR87 (*1)	日本語? (NCR適用による類推)
アラビア語	アラビア文字系の諸言語。ペルシア語、ウルドゥ語など。 (「取扱い及び解説」)	AACR2 (「取扱い及び解説」)	英語 *本文の言語も可能 (「CM」)

(*1)適用する目録規則：「目録情報の基準」による。「原則として日本語資料、中国語資料、韓国・朝鮮語資料については NCR ... 左記以外の資料については AACR2」とある。

また、中国語資料については、「今後の対象言語の拡大を考えると、その都度別の目録規則を必要とするようになることをさけたかったため」に NCR87 を採用したとのことである(宮澤 彰「NACSIS-CAT 総合目録における中国書目録」2001/10/05 <http://www.kanji.zinbun.kyoto-u.ac.jp/courses/2001/comp/NACSIS.txt>)。

(2) タイトルの記述等

	記述に用いる文字	「ヨミ」フィールド	「その他のヨミ」フィールド	ヨミの分かち書き	翻字形の取扱い
中国語	書かれたままの字体 (「CM 案」の例示による)	カナヨミ	ピンイン *任意	行う *漢字キーワード自動切り出しのため	
韓国・朝鮮語	書かれたままの字体 (「取扱い」)	ハングルヨミ (「取扱い」) (*1)	*不使用	行う *漢字・ハングルのキーワード自動切り出しのため	VT に入力可 (*2)
アラビア語	表示されたままの字体等 (「CM」)		LC 翻字形を記録 (「CM」) (*3)	行う (「取扱い及び解説」) (*4)	LC 翻字形を「その他のヨミ」に記録

(*1)ハングルヨミ： 「取扱い」には「必要な場合」とあるが、必須項目と理解できる。

カナヨミ： カナヨミは VT フィールドに入力可となっているが、入力例には記載されていない。タイトルの全てが漢字なら、カナヨミの入力も可能だが、「ハングルのみ」「ハングル及び漢字」の場合、統一的な入力方法を定めるのは困難である。

(*2)韓国・朝鮮語資料の翻字形： VT フィールドに入力可とされているが、翻字方式が複数あり、共同目録 DB である CAT に入力するのは不適當と考えられる。

(*3)アラビア文字資料の翻字形： 議事録によれば、「VT:RM」に入力する案もあったそうであるが、TR フィールドに入力することで、アラビア文字データとの対照性が確保できる。また、今後の多言語運用を考慮し、その他のヨミフィールドに入力するのが適切であると判断した、とある。NII として、その他のヨミフィールドに翻字形を入力することを定着させたいとの意図も感じ取れる。

(*4)ヨミの分かち書き： 各言語の慣習を考慮すると、中国語、韓国・朝鮮語資料は「(分かち書きの習慣がないのに)分けた」、アラビア語は「(当該言語の慣習によりそもそも)分かれていた」と理解できる。

(3) その他

	典拠の取扱い	既存データの取扱い	暫定入力の可否	重複書誌を防ぐための検索時の注意点
中国語	原則として表示されている字体 (*1) (「CM案」)	特に規定されていない。 *実際上は韓国・朝鮮語と同様。	暫定的に漢字を置き換えてもよい。 (オンライン・システム・ニュースレター No.70, 2000.3/17)	特に記載されていない。 *漢字統合インデックスにより概ね無視できる。
韓国・朝鮮語	原則として表示されている字体 (*1) (「取扱い」)	修正可能な参加館が行う。 (「解説」)	不可 (「取扱い」の確定に伴う必然)	暫定入力方式のデータがあることを前提に、漢字・カナ・翻字による検索が必要。 (「解説」)
アラビア語	LC 翻字形を標目形とする。 (「取扱い及び解説」) (*2)	修正可能な参加館が行う。 (「取扱い及び解説」)	不可 (「CM」の確定に伴う必然)	特に記載されていない。 *原綴・翻字による検索は必要と思われる。

(*1)中国語、韓国・朝鮮語資料の典拠の取扱い： 正確には、「原則として最初に典拠レコード作成時に用いた目録対象資料中本体に表示されている字種・字体のままを記載」である。

韓国・朝鮮語の「取扱い」も同じ主旨の記載がある。

(*2)アラビア語資料の典拠の取扱い： 「取扱い及び解説」には、「アラビア文字による標目は、他の諸言語の状況も考慮し、必要に応じて検討」とある。

また、議事録によれば、以下のような検討がされた。

- ・本来は、原綴りによる表記であるべき。
- ・しかし、典拠レコードは全ての目録担当者が使用するものである。
- ・また、多言語対応済のクライアントが少ない。
- ・アラビア文字表記は SF フィールドへ入力する。

4. モンゴル文字資料の入力方針案

当グループでは、3.2での分析に基づき、モンゴル文字資料を現行の環境で行うにはどういった方針が考えられるか検討した。以下に詳細を示す。

検討段階で主に問題となったのは、タイトルの記述に関することであった。

(1) 目録規則等

・ 適用範囲

ウイグル式モンゴル文字で書かれた資料を対象とする。

なお、ここでは主たる情報源にウイグル式モンゴル文字が記載された資料を想定している。

・ 適用する目録規則

AACR2 とする。

「目録情報の基準 第4版」4.1.2の記載によると、日本語資料、中国語資料、韓国・朝鮮語資料はNCR87を適用し、それ以外の言語についてはAACR2を適用する旨の記載がある。

・ 目録用言語

英語とする。

AACR2を適用するため。

(2) タイトルの記述等

・ 記述に用いる文字

ローマ字翻字形とする。

言語の特性により、分かち形で記述する。

本来、転記の原則により資料に書かれたままの字体で記述することが望ましいが、モンゴル文字入力環境が整っていないため、原綴の入力は不可能に近い。奥付の漢字表記の標題を採用する場合でも、AACR2を使う中で違和感が残る。

翻字形を採用することで、将来入力環境が整備された際にも対応が可能となる。

・ 「ヨミ」フィールド

使用しない。

・ 「その他のヨミ」フィールド

使用しない。

- ・ ヨミの分かち書き
使用しない。

- ・ 翻字形の取扱い
LC方式を採用する。
現行の翻字方式で最も一般的と思われるため。

(3) その他

- ・ 典拠の取扱い
著者標目形に関しては、原則として最初に典拠レコードを作成する際に用いた資料に表示されている字体のまま記録する。
モンゴル文字の場合は翻字形を付与し、ヨミは入力しない。
漢字表記の場合は中国語資料と同様に、漢字の標目とカナヨミとピンインを付与する。
他の形の標目はSFに入力し、検索の便宜を図る。
- ・ 既存データの取扱い
各参加館の間で必要に応じて修正を行う。
- ・ 暫定入力の可否
許可しない。
重複書誌の増加につながる恐れがあるため。
- ・ 重複書誌を防ぐための検索時の注意点
約1,500件の既存書誌レコードを考慮し、翻字形、漢字、カナ、ピンインなどで十分な検索を行うことが重要である。

5. まとめ

以上、モンゴル文字の特徴や書誌レコードの入力事例などを整理し、既存の多言語入力規則の要素を比較検討することによって、モンゴル文字資料の簡略な入力方針案の作成に至った。

入力規則というのは本来文字の性質や資料の特性などを専門家の方々に検討していただきながら細かく制定していくものであるが、日々その間に起こりうる目録業務上の問題に鑑みて、それ以前の段階で、入力の指針となり得るものが作成できないだろうかということが今回の当グループの試みであった。

また、当グループでは、本研修のテーマである「品質管理」を、いわゆる多言語資料に関連付けた上で、次のように理解した。それは、(1)書誌レコードの精粗・統一性、(2)書

誌レコードを維持・管理していくための運用上の方向性を定めること、(3)その方向性を各参加組織やその目録担当者等の共通理解としていくこと、そして、(4)重複書誌を増加させないこと、である。当グループが検討した入力規則の方針案は、こういった「品質管理」を考える上で最低限必要なものである。

今回の報告は限られた日数や時間の中で作成したものであり、当グループの4名とも多言語資料についてそれほど詳しいわけではないことも、ここに正直に申し上げる。また、4名のうち2名は多言語「未対応」館からの参加者であった。そのため、それぞれの項目において、誤りや誤解、あるいは不十分な点があるであろうことは、当グループとしても十分に承知している。

一方で、当グループとしては、本研修において、ここで報告した内容を検討することができたことは、たいへんよい機会を与えていただいたと思っている。そして、微力な当グループに、NACSIS - CAT のレコード数の調査や各種会議の資料を提供していただいた、コンテンツ課の岡田様にはこの場を借りて感謝申し上げます。

ここに報告した入力方針案が、今後あるかもしれないモンゴル文字資料の入力方針や方向付けの策定、あるいは現状での NACSIS - CAT への入力の際に、もし少しでもお役に立てれば幸いである。

* 追記

発表時の質疑応答の際、他の言語への適用可能性について質問を受けた。当グループでも他の言語に一般化できたらよいと考えていたが、時間等の制約により今回はモンゴル文字資料についてのみしか検討することができなかった。

この点に関連して、コンテンツ課の川瀬様、岡田様のコメントに基づく議論が今回の報告を発展させる上で示唆に富むものと、当グループは考えたので、以下に要点を整理しておく。

- ・次はどの言語の規則を検討するかと考えた際、選択に困る。また、どの言語まで定めるのが際限がないとも思える。
- ・今回検討した方針を、他のマイナー言語にも一般化できたらよいとの希望もある。
- ・NACSIS -CAT では UCS による入力ということが前提になりつつあるが、UCS に含まれていても入力できないものもある。
- ・文字によって、「翻字グループ」「原綴(UCS)グループ」のように分けるという案も考えられる。