

中国語資料の総合目録データベース（NACISIS - CAT）登録について

岡山大学附属図書館鹿田分館 森谷めぐみ

はじめに

今回学術情報センター（以下センター）で「中国語資料の取り扱い」についての検討案（以下検討案）を目にする機会を得た。センター側では今後、総合目録委員会で検討し、広く公開して意見を募った後、『目録情報の基準 第四版』に反映させていく予定である。長年暫定的なものになっていた中国語資料の登録基準がやっと確定し、図書館の日常的な業務として軌道にのりそうな模様である。総合目録に統合されることで、中国語資料についてあまり知識のない担当者が目録業務、あるいはILL業務の中で取り扱う機会が増えるかもしれない。そこで、中国語資料の特色と、運用に当たって留意すべき点を二三考えていきたい。

1. 中国語資料の古典と新書

中国語資料は大雑把に言って、古典（古籍）と、現在の出版物である新書の二種類に分けられる。前者は伝統的に漢籍と呼ばれ、四部分類など独自の方法で整理されて、十進分類法の適用が可能な後者とは分けられてきた。

古典と新書の分け方はいくつかある。『日本目録規則 1987年版改定版』（以下NCR87）では「辛亥革命以前」という出版年で区別するし、『京都大学人文科学研究所漢籍分類目録』、『東京大学東洋文化研究所漢籍分類目録』などでは内容によって区別される。つまり現代に洋装形式で出版されたものでも、内容が古典なら古典として整理される。中国の国家標準である「古籍著録規則」は、「1911年以前」という出版年と、「古典的装丁」という形態によって区別している。

今回の検討案では古籍の取り扱いについて、「古籍として、別ファイルを設定することはないが、入力規則は別に作成する予定である」とされたので、当分の間総合目録データベース（以下NC）で、はっきりとした取り扱い基準が示されようとしているのは、新書であると考えていい。区別についてセンター側に確認したところ、出版年と装丁による区別を採用し、内容による区別は考えていないということであった。従って現代に出版された影印本のような複製本でも、新書としてNCへの登録が可能になる。その場合はしかし、元々の古典としての性格を考え、巻首書名や元となった版本の記述などを、VTやNOTEなどのフィールドに入れることが利用者にとっては望ましいことであろう。

2. 検討案の基本方針

適用する目録規則は原則としてNCR 87と決められた。ただし、例外的に中国語資料の特性や、今回共に運用が準備されているCHINAMARC¹からの流用入力に対応するため、『中国文献編目規則』の適用をする場合もある。これらの例外的なケースについては、『コーディングマニュアル』等で明示される予定である。

次に今後、文字コードとしてUCS(Universal Multiple-Octet Coded Character Set)²の採用が決まっているが、これにより簡体字の入力が可能になる。このため文字は転記の原則により、書かれたままの字体で記録する。実際にはこれまでの繁体字でのデータの蓄積もあり、繁体字と簡体字が混在することになるので、異なる字体間での一括した検索を可能とするため、漢字統合インデックスが作成される。

最後にヨミの取り扱いであるが、漢字の単語単位での検索を可能とするため日本語ヨミの付与が必須とされた。それに伴い、中国語資料の日本語ヨミと分かち書きの規則が新規に作成される。(これは検索のための便宜的な規則であり、センター側で中国語の日本語ヨミの標準化を定めるというのではない)

本来外国語である中国語資料を取り扱うに当たって一番困難で運用が揺れると思われるのは、この翻訳の問題とも関係する日本語ヨミと分かち書きの部分ではないだろうか？ 今回この規則を検討案の中で目を通させてもらったが、かなり大部にわたる複雑なもので、中国語の文法的な知識を要求される部分もあり、これを無理なく運用するにはかなりの習熟が必要になるという印象を持った。センター側に訊いたところ、日本語ヨミの付与はあくまで漢字による検索キーを切り出すためのものであり、ある程度の揺れは覚悟している、検索は主に漢字キーを使ってほしいということであった。

もうひとつ有効な検索キーになりそうなものにピンインがあげられるが、入力業務担当者の負担を考え、今回これは選択事項とされた。最もCHINAMARCではピンインは付与されているのでそこからの流用入力により入ってくると考えられ、ある程度検索キーとしての機能を果たすものと考えられる。ピンインの分かち書きに関しては、CHINAMARCに準拠し、漢字一字ごとに分かちされ、単語ごとの分かちはされない。そうなれば、分かちの揺れを考えなくてもよいし、ピンインは一つの漢字に一つの音が割り当てられることが大部分なので、より正確な検索が行える。また新CATのフルタイトルキーによる検索機能を使えば、語順の指定もできるので、検索ノイズが増える心配もない。入力率次第によって、効果的な検索キーになるのではないかと思う。

3. 雑誌目録について

雑誌目録はILLでもよく利用されるため、検索の面から少しここで考えてみたい。

欧米のデータベースや資料からの引用文献で申し込まれる場合、たいていそれは漢字ではなく、ピンインか翻訳された英文名で記述されている。そこで国内の目録で所蔵館を探す場合、漢字への翻訳作業や、中文誌名と英文誌名の対照表などに当たる必要がある。

NCはどうであろうか？ 現在NCの雑誌書誌は、ピンインや英文ではなく漢字だけで書名が登録されているものも多いので、ピンインで検索するとNCを飛び越してLCにヒットしてしまうこともあり、注意が必要である。検索のヒット率を上げるにはどうすればいいだろうか？ ピンインの入力が効果的と考えられるが、2.でも書いたように今回の検討案ではピンインの分かち書きはCHINAMARCに準拠し、漢字一字ごととされている。引用文献によっては単語ごとの分かち書きで表記されることも考えられるので、書かれたままの形で入力しても、ヒットしないケースが出てくる。漢字一字ごとに分解すると言っても、どこで切っているのか、中国語の知識のない者にとっては分かりにくい。となれば書誌の方で、ピンインを漢字一字ごとに入れた上、検索用として単語ごとの分かち書きをVTにでも入れるべきだろうか？ 検索の面から言えば理想的だが、入力作業を増やすことになるので、あまり普及するとは思えない。

検索面から考えるとやはり有効なのはISSNだろうか？ 二次資料を検索し直してISSNを確認し直すか、中国語資料をよく利用する利用者には申し込みの段階からISSNを書いてもらうよう、呼びかけることが必要になってくるだろう。最もISSNも全てのデータについて入っていると限らないので、検索漏れの可能性がある。どちらにしても中国語雑誌の検索については、上記のような事情があるので、複数の検索キーを使い分けるよう、ILL業務担当者に注意を呼びかけることが必要になってくるだろう。

4. 今後の予定

さて中国語資料の登録に当たって今後の予定であるが、センター側では今年11月には検討案を公開し、意見を募ることになっている。またUCS使用可能なサーバの公開が2000年1月に予定されている。クライアント側の対応として、中国語入力が可能な端末が必要になるが、今のところUCSの使用可能なクライアント側の機器が現れていないので、現実的な方法としてWindows 95の中国語版であるChina Windowsが考えられている。これにより中国の国内規格であるGBコードを使った簡体字の入力が可能になる。

クライアント側でGBコード、あるいは従来通りのJISコードで入力されたデータはセンター側のサーバでUCSに変換される。これを各端末側で、扱うことのできる文字コードに再度変換し読み取ってくるわけだが(表1.)、そのコードの中にUCSに対応する文字が準備されていない場合、現在の外字のようにマークと共にUCSのコードが表示されることになる。例えば、日本のWindows 95ではGBコードにしかない簡体字は、文字ではなくコードの形で表示されるし、逆にChina Windowsから繁体字(GBコードにはない)である日本、台湾の漢字を見てもコードの形で表示される。

登録する側から言えば、JISにより登録を行う館は、簡体字を外字としてコードの形で入力する方法もあるだろう。どちらにしても面倒な話だが、いずれはWindows NTなどで、UCSの使用が可能になり、日本語、中国語、それぞれの文字コードの文字数の差を考えなく

ても、UCSとして一括して処理できるようになることが期待される。それまでの暫定的な処置として考えたい。

5. おわりに

今回の検討案は細かい部分はまだ変更があると思うが、基本方針についてはほぼこの方向で行くであろう。この流れは新CATの多言語システムへの対応の一環とされるものであり、次の段階としては朝鮮語資料などへの対応も予定されている。

多言語資料の総合目録化というのは、資料の幅広い公開という点では意義のあることだが、言語についても資料についての専門的知識のない者が、業務の中でいきなりそれに触れる機会が増えることも意味する。混乱を防ぐための知識の普及や基礎的な語学の勉強、経験の少ない担当者のためのわかりやすいマニュアルの作成などの準備を、目録作成館の側で今から進めておく必要があるだろう。

<参考文献>

- * 高橋菜奈子「中国語資料のISBNについて：UCS導入後をめざして、書誌作成館は何に留意しておくべきか」(平成9年度第2回総合目録データベース実務研修レポート URL：<http://www.nacsis.ac.jp/hrd/cgi-bin/viewdoc.db2.pl?ACTION=VIEWPAGE&DBNAME=1997.db&ID=db.1997.atl20&PAGE=1>)
- * 松本浩一「標準化・機械化の中での漢籍目録」(『漢籍 整理と研究』No. 5 p. 15 - 24, 1995年)
- * 安達勉「漢籍研究会1994年度夏季研修会研究発表要旨」(『漢籍 整理と研究』No. 5 p. 11 - 14, 1995年)

¹ 中華人民共和国の国立図書館である北京図書館が作成。台湾の図書については入っていない。ところで、雑誌の書誌については入っているはずなのだが、現在のところその存在が確認できていない模様である。雑誌についてCHINAMARCからの流用入力ができないとなると、全くの新規登録か、LCからの流用入力のどちらかで対応しなければならないだろう。数的には限られているかもしれないが、今のところ中国語雑誌の登録は図書よりは煩雑になりそうである。

² 日本語、中国語、ハングルなども取り扱えるユニコードの文字セット。1993年に国際規格ISO/IEC 10646-1として制定され、1995年1月にJIS X 0221として日本の規格に採用された。

表1 . サーバ、クライアント間での文字コードのやり取り

