10-1-2002

The Pinyin Conversion Project of the University of Michigan Asia Library

Mei-ying Lin
Fang-hwa Chen

Follow this and additional works at: https://scholarsarchive.byu.edu/jeal

BYU ScholarsArchive Citation
Available at: https://scholarsarchive.byu.edu/jeal/vol2002/iss128/4

This Article is brought to you for free and open access by the All Journals at BYU ScholarsArchive. It has been accepted for inclusion in Journal of East Asian Libraries by an authorized editor of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen amatangelo@byu.edu.
THE PINYIN CONVERSION PROJECT
OF THE UNIVERSITY OF MICHIGAN ASIA LIBRARY

Mei-ying Lin and Fang-hwa Chen
University of Michigan

Background

The Asia Library of the University of Michigan is one of the major comprehensive research collections of Asian materials in North America. The collection has grown from less than 20,000 volumes in the early 1950s to 714,664 volumes/reels/sheets at the end of fiscal year 2001-2002. In addition, there are several hundred CD-ROMs and video tapes. The Chinese collection consists of 351,240 volumes, 29,563 microfilm reels, 24,115 microfiche sheets, and more than 400 CD-ROMs.

To provide bibliographical access to the collections, the Asia Library published its twenty-five volume printed catalog, Catalogs of the Asia Library (Boston: G. K. Hall, 1978). We began cataloging in RLIN-CJK in 1984 and in 1991 completed the retrospective conversion of all bibliographical records into machine-readable data.

The Pinyin Conversion Project

In transliterating the Chinese elements in bibliographic records, the Wade-Giles Romanization system had been standard in American libraries from the first days of Chinese collections. On February 11, 1958, the Chinese Government promulgated “Hanyu pinyin fangan” (Plan for phonetic writing of Chinese). Hanyu pinyin is used for teaching Chinese in China and was introduced to the international community at 1978 UN Conference on Geographic Names. In the years following 1978, Pinyin gradually gained acceptance in American universities and in Western language publications as the standard to transliterate Chinese. American libraries recognized the necessity to change and on October 1, 2000, began using Pinyin to transliterate all Chinese elements in bibliographical records. In order to provide better service to users of the Asia Library and help them use our collection more efficiently, we planned the Pinyin Conversion Project with realistic assessments and made decisions that would most facilitate usage of the library.

I. Planning

In May, 2000, the Head of Asia Library, the Coordinator of Information and Public Services, the Coordinator of Technical Services, and the Pinyin Task Master discussed the issues and assessed the situation. The discussion was summarized in the “Pinyin Conversion Project Studies and Planning,” in which we identified the scope of the project and outlined steps to convert the 160,000-200,000 records.
II. Testing

From the end of May 2000, the Pinyin Task Master participated in the RLG Pinyin Working Group conference calls. In the conference calls, the working group members discussed the issues related to the Pinyin Conversion Project. Usually the conference calls lasted one and an half hours to two hours.

At the beginning of June 2000, our library decided to join the RLG testing group with six other libraries: Columbia University, Princeton University, Yale University, Cleveland Museum of Art, University of Toronto, and the University of Washington Gallagher Law Library. We were requested to submit fifty of our records to be tested. We selected 50 records (38 monographs, 7 serials, 4 microfilms, and 1 map) to be tested and submitted them to RLG on June 15, 2000.

When the test was released, the Pinyin Task Master reviewed the test first, then submitted the comments to the RLG, and discussed the results with other testing groups in the conference calls.

The following are the results of the tests:

1\textsuperscript{st} test
In the first release, the programs for personal and geographical names were not yet fully in place. Taiwan place names were converted when they should not have been. Also, some fields, such as 501 (with note), 780 (preceeding entry) and 785 (succeeding entry), were not converted. Some other fields and some words were not converted. The test reviewers requested the addition of field 987 (local romanization/conversion history) in every record.

2\textsuperscript{nd} test
In the second release, improvements were made, but some problems still remained. Again, personal and geographic names were not yet fully in place. The problem of Taiwan place names was not solved. Some mixed texts were still not converted completely. However, we did see some improvements in the second test. Field 987 was shown in every record, although some subfields were still under discussion. In our own records, field 501 was converted successfully. We were pleased to see that field 501 was converted because many of our series analytics contain this field.

3\textsuperscript{rd} test
In the third release, the issue of personal and geographical names was still a major topic. The test mainly focused on the “mixed text.” Some fields were defined to be handled as “mixed text.” The test reviewers also agreed to add subfield “f” to field 987 to indicate which fields were not converted and required manual review.

4\textsuperscript{th} test
In the fourth release, personal and geographic names worked fine. Some Taiwan place names were converted successfully. Participants were happy to see the field 987 with
subfield “f.” We also learned that family names in field 600 (subject added entry—
personal name), such as “Chang family, Ch’en family” would not be converted at this
time. Libraries will wait for LC’s decision on this issue.

5th test
The 5th test was released on September 25, 2000 (far behind schedule). In this release,
more information was found in field 987 subfield “f” to indicate which field needed
manual reviewing. No further conference call was arranged to discuss the 5th test. For
the five tests, we saw some improvements in every test. However, we still faced some
problems, such as the “mixed text” issues and typographical errors. For those records,
we needed to do manual conversion.

III. New Classification Schedules

The Classes “DS” (History of Asia) and “PL” (Language of Eastern Asia) would be
affected most by the project. The Library of Congress proposed a new “Classification
Schedule for Local History and Description of China” that was somewhat different from
what we had been using. In order to be consistent with Library of Congress practice, we
decided to follow the new schedule. But we had to find a way to deal with the existing
collection because the call numbers of conventional place names and personal names
could be changed by the new schedule and new Cutter numbers. We had to decide either:

(1) Apply the new schedule and rule to every record, which would mean we would
have to change the call numbers of all involved records in our library collection.
By doing that, our records would be consistent with the new classification
schedule and rule. However, a great deal of manpower and time would be
necessary to finish the job.

(2) Apply the new schedule and rule from October 1, 2000, leaving the call numbers
of existing records as they were. That would mean split files: the same places and
personal names could be shelved under different call numbers. In order to
minimize user inconvenience, we could make signs in the stacks to guide patrons
to find materials with both the old numbers and new numbers.

After discussion among those librarians responsible for Chinese materials, we made the
following decisions:

(1) For the DS (History of Asia) section, we decided to follow the new
“Classification Schedule for Local History and Description of China” proposed
by the Library of Congress. In the new schedule, call numbers for provinces are
not changed dramatically, however, call number for cities are modified. In the
old schedule, cities were arranged alphabetically regardless of their location. In
the new schedule, cities in the same province are arranged together in
alphabetical order. This is an improvement.
(2) For the PL (Language of Eastern Asia) section and individual works and biographies, such as Mao Zedong, Deng Xiaoping, we chose to use the current numbers, instead of creating new Cutter numbers. This will avoid patrons having to find materials under two different call numbers.

IV. Cataloging and Authority Files

We determined to use Pinyin to transcribe all Chinese bibliographic records beginning from October 1, 2000. Field 987 (Local romanization/conversion history) would be applied to the bibliographic records as needed. By that date RLG had finished loading the converted authority records into RLIN NAF file, and our plan was to follow the authority file records as changed. For MIRLYN (the Local system), we received our authority records from RLG. Basically, we would keep the same rules as ordinarily used for cataloging except that certain Pinyin syllables would be joined together based on the guidelines issued by the U.S. Board on Geographic Names. For example, we used “Chung-hua jen min kung ho kuo” before; now it would be “Zhonghua Renmin Gongheguo”. Since patrons searching MIRLYN to locate relevant materials might have trouble finding what they wanted because of the change, we planned to mention this change in our user education efforts.

It was decided that in the intervening time before the conversion Chinese catalogers would do copy cataloging with Wade-Giles records so that records in MIRLYN would not be mixed Wade-Giles and Pinyin. We continued doing this until Wade-Giles records in MIRLYN were all converted to Pinyin. Thus, we managed to avoid creating split files in the local system.

V. Reference Materials

For help in applying the new transliteration system for Chinese materials, we selected some standard materials as our references. First, we would use the “Library of Congress Pinyin Conversion Project—New Chinese Romanization Guidelines—Rule of Application & Correspondence of Wade-Giles to Pinyin”. Second, we would use some standard dictionaries as our reference tools.

165,000 Records Converted

In order to avoid creating split files and facilitate searching, it was necessary to convert all Chinese bibliographical elements and headings into Pinyin. With the support of the Library Administration, Monograph Cataloging, Serials, Authorities, and the Library Systems, 164,984 Asia Library Chinese records were converted and received from Research Libraries Group by April 30, 2001. The Asia Library not only successfully implemented a large and complex project but also managed to avoid the confusion and inconvenience that commonly accompany such projects.
Manual Review Completed

In cooperation with the Library of Congress and the Council on East Asian Libraries, the Research Libraries Group also screened the Chinese records to identify those that needed manual review. The 28,622 records so identified, 17% of the Asia Library Chinese records in RLIN, were printed by the System Office from the Snapshots received from RLG. The records are categorized as follows:

<table>
<thead>
<tr>
<th>Fields</th>
<th>Numbers</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1XX</td>
<td>264</td>
<td>MIUOAHF9052-B</td>
</tr>
<tr>
<td>2XX</td>
<td>6648</td>
<td>MIUO87-B7156</td>
</tr>
<tr>
<td>4XX</td>
<td>1002</td>
<td>MIUO91-B4601</td>
</tr>
<tr>
<td>5XX</td>
<td>15714</td>
<td>MIUO98-B5545</td>
</tr>
<tr>
<td>6XX</td>
<td>255</td>
<td>MIUO00-B4206</td>
</tr>
<tr>
<td>7XX</td>
<td>893</td>
<td>MIUO84-B2279</td>
</tr>
<tr>
<td>8XX</td>
<td>231</td>
<td>MIUO98-B6443</td>
</tr>
<tr>
<td>Not converted</td>
<td>202</td>
<td>MIUO91-B4731</td>
</tr>
<tr>
<td>Others</td>
<td>3413</td>
<td>MIUO88-B2389</td>
</tr>
</tbody>
</table>

Manual review of records converted by RLG began at the Asia Library on May 18, 2001. We used the printouts as worksheets and updated the records one-by-one on RLIN. From May 2001, two librarians and two student assistants worked continuously on this project, and by March 31, 2002, all 28,622 records had been reviewed, revised, or updated. The large number of records flagged for manual review reflects the complexity of the Project rather than the quality of our records. For instance:

(1) Some records that were flagged should not have been changed, e.g. some personal names and Taiwan place names. (MIUO89-B3267)

(2) Typographic errors were responsible for many records marked for manual review. Most of those records were easily figured out and corrected. However, there were some unique Asia Library records in RLIN. For these, we first checked the card catalog and then updated the records. (MIUO85-B3927)

(3) Some records were not converted appropriately. For some serial records, “no” (number) is converted “nuo.” (MIUOAHG5307-B)
(4) Mixed-text. Some mixed-text subfields marked for review were partially converted (MIU084-B2279)

The field most frequently marked for review was 5XX. In the conversion of most of the other Asian libraries’ records, the 245 field was the field most frequently identified for review. Our distinctiveness was due to our Chinese Collectanea Analytic Project. We had been converting records for the titles in this project since 1997. All the analytics contained a 5XX field, and many were not converted correctly.

**Concluding Remarks**

As we complete this large and complex project, it is gratifying to note that the user is better served because of the greater degree of standardization. The standardization of the transliteration system and headings for Chinese, for example Tianjin, Sichuan, Shaanxi, has eliminated some of the ambiguities most often causing complaints. It will also help facilitate international exchange of bibliographical data.

**References**


(3) Summary report of the CEAL task force to review a possible change from the Wade-Giles to the Pinyin romanization system. *Journal of East Asian Libraries*. No. 115 (June 1998):40-44.


