10-1-1998

Pinyin Conversion Reports

Robert Hiatt

Philip Melzer

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

BYU ScholarsArchive Citation

This Report 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.
PINYIN CONVERSION REPORTS

Joint CEAL Task Force on Pinyin/LC/OCLC/RLG Meeting

March 25, 1998

Attendees: Randall Barry (Library of Congress), John Byrum (Library of Congress), Timothy Connor (Harvard University), Martin Heijdra (Princeton University), Robert Hiatt (Library of Congress), Karl Kahler (University of Pennsylvania), Karl Lo (University of California, San Diego), Tai-loi Ma (University of Hong Kong), Philip Melzer (Library of Congress), Beatrice Ohta (Library of Congress), Anthony Pierce (Library of Congress), Karen Smith-Yoshimura (Research Libraries Group), Amy Tsiang (University of California, Los Angeles), Cynthia Whitacre (OCLC), Beacher Wiggins (Library of Congress), Thompson Yee (Library of Congress), Peter Zhou (University of Pittsburgh)

Introduction: Melzer welcomed the attendees to the meeting and briefly reported on the activities of the Library of Congress, particularly the Task Group on Pinyin Conversion which had been meeting regularly over the past several months. He suggested that Oct. 1, 2000, be penciled in as a target date, although he noted that in his conversations with OCLC this date may be ambitious for OCLC. It is hoped that by that date RLG can convert Chinese-language records from Wade-Giles to pinyin romanization.

Conversion: Smith-Yoshimura noted that RLG had offered to convert all Chinese-language records in RLIN from Wade-Giles to pinyin in a response to a letter from Wiggins. She also noted that not all Chinese-language records residing in the RLIN database have Chinese script; some institutions input only the romanization. The cost of the conversion will be absorbed by RLG, but each institution will be charged standard rates to receive a copy of its own converted records for loading into its local system.

Melzer noted that in the conversion process, only “pure” Wade-Giles strings would be converted. Strings based on Wade-Giles romanization or with combinations of Wade-Giles and “conventional” romanizations would not be converted.

Guidelines

Aggregation: Melzer noted that the Library of Congress’ recommendation will be not to aggregate syllables. This recommendation is based upon the requirements of those institutions that have followed LC’s past practices. The only aggregation of syllables (word division) that LC will recommend will be for personal forenames and geographic names. Lo noted that the CEAL task force strongly endorsed word division following the PRC standard. He said that the compelling reason for conversion is to adhere to the Chinese standards, and it would be counterproductive to deviate from those standards, particularly on word division, i.e., zhengcifa. He noted that syllables aggregated into words can always be divided into syllables, but the reverse is not true. Kahler felt that the plans should be developed for the future and have the old fit into the new structure, not vice versa. Lo noted that traditionally aggregation (as word division) has been applied to Japanese- and Korean-language materials as part of the romanization tables. Melzer noted that for Chinese-language materials, it is often difficult to
determine the “word.” Smith-Yoshimura noted that if the romanization were not aggregated, then presumably neither would the vernacular script be aggregated, and there is significant value to researchers to search across Chinese-, Japanese-, and Korean-language material by using CJK terms common to all three.

Di/de: With respect to the romanization di/de, Melzer noted that although the National Library of Australia reviewed each usage in the records it converted, the Library of Congress will recommend a continuation of current policy and always use “di” (“ti” currently). Lo noted that the PRC standard is to use “de.” Melzer noted that the conversion will result in “di” which may need to be changed later.

Wade-Giles title proper: In response to Melzer’s question concerning the necessity to retain the Wade-Giles title proper in a 246 field in the converted records, the response of those in attendance was negative.

Subject headings: Melzer noted that a task group has been established at the Library of Congress to consider subject headings in Wade-Giles romanization or those based on Wade-Giles romanization.

Geographic names: Hiatt detailed his recommendations to the Library’s task group on the handling of jurisdictions and populated places in China (including Hong Kong) and Taiwan. The recommendations are 1) choose the forma approved by the United States Board on Geographic Names with preference given to any brief, conventional form in English; 2) if no BGN form is found, choose a romanized form found in the item; 3) if there is no found-romanization, use the systematically romanized (pinyin) form. Hiatt noted that although no decision had been taken, there has been discussion of separating the change in geographic headings from the general conversion to pinyin. Because geographic names are used as qualifiers to other headings, this would result in a mixture of romanizations in a character string. Heijdra noted that in some cases a two-step conversion may be necessary.

Time frames: Melzer asked if there were any thoughts on the time frame involved with the actual conversion of individual records. The longer the time frame, the longer the institutions will be dealing with split files. Both OCLC and RLG indicated that the effort to reach full conversion will amount to months rather than years. Smith-Yoshimura felt that establishment and adherence to a tight time frame for conversion activities would prove to be cost-effective.

Melzer noted that the Library of Congress would be working with RLG to draft a conversion time line.

Classification: Melzer noted that the Library of Congress will need to simplify the conversion as much as possible and be as cost-effective as possible. For this reason the Library will not reclassify material when its alphabetical arrangement within the classification schedules changes because of a change in entry element (e.g., Beijing vs. Peking). Yee noted that this policy, which was followed in the change to AACR2, refers to those cutters printed in the schedules, not to those that are not printed.

Future cooperation: Both Lo and Melzer hoped that CEAL would appoint a representative
person or group to continue its liaison with the Library of Congress. Smith-Yoshimura suggested that a report be prepared highlighting questions and issues that were not addressed in this meeting.  

(Robert Hiatt, Library of Congress)

**Pinyin Conversion**

**NOTES OF A MEETING HELD ON JUNE 27, 1998, 4:00-6:00,**  
Sheraton City Centre, Washington, DC

**Attending:**

An Shulan, Director, Academic Committee, Tsinghua University Library, Beijing

Nancy Ou-lan (Hu) Chou, Director of Libraries, National Chengchi University, Taipei

Yu-lan Chou, Cataloger, East Asian Library, University of California-Berkeley

Hwa-Wei Lee, Dean of University Libraries, Ohio University

Philip Melzer, Team Leader, Regional and Cooperative Cataloging Division, Library of Congress

Beatrice Ohta, Team Leader, Regional and Cooperative Cataloging Division, Library of Congress

Sun Liping, Director, International Cooperation Division, National Library of China, Beijing

Zhou Heping, Deputy Director, National Library of China, Beijing

Yu-lan Chou and Dr. Hwa-Wei Lee were thanked for arranging this meeting between representatives of the National Library of China (NLC) and the Library of Congress (LC) to exchange information and views about pinyin romanization. Philip Melzer provided a status report on Library of Congress efforts to convert to pinyin romanization. Although LC has wished to convert for a number of years, it was only after the National Library of Australia converted its Chinese bibliographic records that the decision was made to proceed. The two major national utilities, RLG and OCLC, have agreed to work with LC on a national conversion project.

The conversion project was undertaken in 1997 with the decision of LC to convert. A new Chinese romanization standard, based on pinyin, will be issued in the near future. LC has decided that the guidelines will continue the practice of dividing individual syllables with a space, except in personal names and geographic locations. Under the new guidelines, tones will not be indicated, and words of non-Chinese origin will be romanized systematically. Draft romanization guidelines were distributed, and comments were invited.
After conversion specifications have been drawn up, a program will be written and tested. Most portions of bibliographic records will be converted by computer program; subject headings may be converted manually. During conversion, a list of access points will be compiled. Then, the name authority records covering those access points will be converted. At this time, the method of converting name authority records has not been decided upon. The final step in the project will involve converting access points on non-Chinese bibliographic records to assure consistency in the files.

Although the works of literary authors will not be reclassified, pinyin cross-references will be added to the approximately 11,000 Chinese names in the classification schedules to lead users from the new form to the proper class number. Because the spelling of most place names will change, material about a place may be shelved in more than one location in the future. Examples were distributed showing corresponding bibliographic and authority records, and a section of the classification schedule, in both Wade-Giles and pinyin romanization.

At about the time of conversion of bibliographic records, and presumably sometime in the spring or summer of the year 2000, "Day 1" will be declared, after which time all romanization of Chinese will take place in pinyin. However, because LC is installing a new Integrated Library System, there could be delays in implementation.

Sun Liping conveyed responses to LC questions about NLC uses of pinyin romanization, and requests for advice. NLC uses pinyin romanization in bibliographic records to provide another avenue of access (in addition to the vernacular) to title and author. The Chinese MARC record used two subfields for only that purpose. For example, in the 200 field (the equivalent of the USMARC 245 field), subfield a gives the title, followed by subfield A for the pinyin equivalent; subfield f gives the statement of responsibility, followed by subfield F for the pinyin equivalent.

Romanization is done by computer program, and then given a human check for accuracy. Each syllable is separated, and tones are not indicated. If a term has been transliterated from a foreign language into Chinese sounds (i.e. Kekoukela), then that term is not romanized again back into pinyin, but translated words are romanized systematically.

Ms. An was supportive of the LC decision to generally separate syllables. She felt that Chinese aggregation guidelines were problematic and not workable for purposes of indexing; they are intended to be applied to the modern language, but not to classical texts. She felt that connecting syllables in personal names and geographic locations would help eliminate ambiguities; her only concern was for the occasional problematic exception (Xi'an and Xian). Ms. An said that adding tones to romanized syllables would provide further indexing to pinyin syllables. She suggested using tones for filing purposes, but not for display.

Ms. Sun noted that some computers in China have limited character capability; therefore, searchable romanized fields are useful in providing users with additional access points. Romanized searches are made by keying in roman letters, and conducted by a different search engine than vernacular searches. Compressed key title searches are based on LC specs for MUMS: 4,3,1 and 3,1,1,1.
Ms. Sun said that NLC routinely uses three dictionaries to determine contemporary pronunciation:

- *Xian dai Han yu ci dian*
- *Zhonghua ta zi dian*
- *Han yu da ci dian*

Ms. Sun said that *Ci hai* is generally useful for older terminology.

Ms. Sun requested that LC send an outline of its conversion plans to her. She in turn would ask experts at NLC to respond and offer advice. She also invited LC representatives to visit NLC to explore ways to facilitate the exchange of bibliographic data between the two libraries. The meeting closed with general discussion and expressions of gratitude.

(Philip Melzer, Library of Congress)

**Brief Update on Pinyin Conversion**

Pinyin romanization guidelines are being finalized, and will be disseminated in the near future.

The Research Libraries Group (RLG) has announced that it will work with the Library to convert its Chinese records to pinyin. In consultation with RLG, the Library has proposed an implementation timeline for certain portions of the conversion project. It is now anticipated that Chinese records will be converted in the spring or summer of the year 2000. Library staff has begun to draw up conversion specifications which will be based on the new romanization guidelines. After specifications have been agreed upon, the Library will then work with RLG to prepare a conversion program.

A subcommittee of the Library's Pinyin Task Group is trying to gauge the scope and impact of conversion on subject headings and classification schemes. The group is investigating how Wade-Giles terms in subject headings can be located in the online catalog, how to decide which terminology should be converted to pinyin forms, and when the Library should initiate changes to subject headings and authorities. A number of changes in the DS schedule are being anticipated. Classification schedules for Chinese literary authors will be expanded to accommodate the new surnames which will be established according to pinyin romanization. References will be provided in the schedule from pinyin forms of name for literary authors. Conversion of subject headings could begin before the year 2000.

Decisions on conversion of authority records will be made after more is known about the capabilities of the Library's new Integrated Library System (ILS).

(Philip Melzer)

**Headings for Chinese Conventional Place Names**

The Library of Congress has reevaluated the headings for Chinese place names that are not entered according to standard Wade-Giles form, as well as certain systematically romanized headings. The Library circulated the proposed changes in the spring and received comments.
from CEAL as well as individual librarians. Most of the conventional headings for Chinese place names listed in Library of Congress Rule Interpretation (LCRI) 23.2, and several others, will be changed to current forms based on pinyin romanization. LCRI 23.2 will be revised to reflect these changes.

The revised LCRI 23.2, along with a list showing both the current forms of headings for Chinese conventional place names, and the new forms to which they will be changed, will be published in Cataloging Service Bulletin (no. 82).

The Library has undertaken a pilot project to determine the most efficient and effective means of making these changes to bibliographic and name authority records. During the pilot project, headings for several place names in Sichuan Sheng (Szechwan Province) will be changed, along with the heading for the province itself:

<table>
<thead>
<tr>
<th>LCCN</th>
<th>OLD HEADING</th>
<th>NEW HEADING</th>
</tr>
</thead>
<tbody>
<tr>
<td>n81-20323</td>
<td>Szechwan Province (China)</td>
<td>Sichuan Sheng (China)</td>
</tr>
<tr>
<td>n81-20054</td>
<td>Ch'eng-tu (China)</td>
<td>Chengdu (China)</td>
</tr>
<tr>
<td>n81-82039</td>
<td>Ch'ung-ch'ing shih (China)</td>
<td>Chongqing (China)</td>
</tr>
<tr>
<td>n89-125551</td>
<td>Hsi-k'ang sheng (China)</td>
<td>Xikang Sheng (China)</td>
</tr>
<tr>
<td>nr93-5302</td>
<td>Zamtang (China : District)</td>
<td>Zamtang Xian (China)</td>
</tr>
<tr>
<td>nr94-43623</td>
<td>Zoige (China : District)</td>
<td>Zoige Xian (China)</td>
</tr>
<tr>
<td>nr94-43605</td>
<td>Barkam Xian (China : District)</td>
<td>Barkam Xian (China)</td>
</tr>
</tbody>
</table>

Because the many changes to authority and bibliographic records will be made incrementally, the Library’s database will, for a period of time, contain headings both in the current forms and the new forms. Before the end of the year, the Library will announce implementation targets.

(Philip Melzer)