Summary of the Survey on *Pinyin* Romanization

Amy Ching-Fen Tsiang
SUMMARY OF THE SURVEY ON PINYIN ROMANIZATION

Prepared by Amy Ching-Fen Tsiang

The Library of Congress (LC) is again exploring the feasibility of converting from Wade-Giles to pinyin romanization. In 1990, LC decided that, prior to any possible conversion, there would have to be agreement on a pinyin word division scheme. With that in mind, LC put out a recommendation concerning pinyin word division as a preliminary step in the process. [The LC proposal is appended to this report--Ed.] In its proposal, LC recommended following the National Library of Australia (NLA) pinyin word division practice, which separates individual Chinese characters from each other. The CEAL Technical Processing Committee was requested to respond to LC’s recommendation on behalf of the profession, hence the survey on pinyin romanization was conducted.

The survey, for the most part, was designed to solicit opinions and comments on LC’s word division proposal. Therefore, this summary of the survey focuses on the responses to LC’s proposal, and should not be considered a thorough discussion of the issue of pinyin romanization.

To conduct the survey, a questionnaire was distributed to 80 East Asian collections in North America. A total of 67 responses, including 65 returned questionnaires and 2 emails, were received. Among them, 17 represented the views of institutions and 50 were individual opinions (including 40 library professionals and 10 teaching faculty members). Following is a summary and statistical report of the survey.

Feasibility of Conversion from Wade-Giles to Pinyin

Of the 67 responses to the survey, 56 (84%) agreed that Wade-Giles romanization should be converted to pinyin. Six respondents (9%) opposed the idea, and five (7%) had no comment. All ten faculty members supported the conversion.

Reasons given for supporting conversion from Wade-Giles to pinyin romanization were

1. *Pinyin* is a standard romanization system. It is a national standard maintained by a national office in China. In contrast, Wade-Giles is no longer a standard that anyone, including the Library of Congress, is willing to maintain. While we are without a "standard bearer," it is now time to change.

2. Pinyin romanization is a far more commonly-known and widely-used system in education related to China studies in North America, and it is also a more accurate romanization system for Chinese pronunciations. In contrast, Wade-Giles is no longer taught in regular schools worldwide. It is almost a dead system that exists merely in library catalogs.

3. The *pinyin* system is used by the majority of Chinese-speaking people, including 1.2 billion Chinese in the PRC, and it is officially accepted by the United Nations. In contrast
to this, Wade-Giles is a system without roots among native Chinese-speaking people.

(4) *Pinyin* is a simpler system than Wade-Giles.

In general, those opposed to conversion do not deny the advantages of *pinyin* romanization. The complexity and cost of conversion were rather the major concerns.

(1) The cost and time input associated with implementation must be taken into account. It is necessary to have a thorough discussion on possible sources of funding, maintenance of existing Wade-Giles bibliographic records and Name Authority file and other issues before starting any conversion project.

(2) Technically, it is possible to develop software that will display Wade-Giles records in *pinyin*. If it is workable, it would not be necessary to convert from Wade-Giles to *pinyin* romanization.

(3) Displaying vernacular characters is more important than converting from one romanization system to another. So long as the online record can display vernacular characters, it does not matter whether the romanization used is Wade-Giles or *pinyin*.

(4) Neither Wade-Giles nor *pinyin* are helpful to non-Chinese-speaking people in pronouncing Chinese.

Many supporters of conversion agree that further discussion and investigation is necessary. They also mentioned that, from a long term point of view, an integrated and standardized romanization system could cut the cost.

**Library of Congress Word Division Proposal**

There were two parts to this section. Many respondents omitted the second part. Of the six responses opposed to conversion from Wade-Giles to *pinyin* romanization, four did not continue the questionnaire.

1. **Do you concur with these aspects of the Library of Congress Proposal?**

(A) that, generally, syllables would be divided from each other

52 of 63 total responses (82%) supported this proposal, 10 (16%) opposed it, and 1 (2%) had no comment. There was a significant difference in the responses of library professionals and teaching faculty members. Of the ten responses opposing this proposal, seven were from faculty members. 92% of librarians supported the proposal, while only 30% of the faculty members did.

Reasons given for supporting the above aspect of the proposal were
(1) Ambiguities in word division will be minimized. It is how Chinese write Chinese. Word aggregation is more meaningful for reading, but it is prescriptive, not descriptive.

(2) The National Library of Australia has already begun using this practice. Following their lead as a standard seems better than being unique.

(3) Having syllables divided would potentially make converting existing Wade-Giles records into pinyin by automation easier.

(4) This practice would eliminate complicated sets and subsets of rules. It could reduce troubles and mistakes in combining characters into words such as presently occurs in romanization of Japanese and Korean.

Opposing views base their opinion on the following reasons:

(1) Meaning in Chinese is based on words, not single characters or syllables.

(2) Aggregating syllables reduces the trouble caused by homonymous characters.

(B) that, only the syllables of surnames, as well as geographic names, would be connected

Once again, 52 of 63 total responses (82%) support this proposal, 10 (16%) oppose it, and 1 (2%) had no comment. The same disparity between teaching faculty and librarians noted in the previous response also was true here: of the ten opposing responses, seven were from faculty members. 94% of librarians supported the proposal; only 30% of faculty members did.

The main reasons for supporting and opposing this proposal were similar to (A) above. Some supporting opinions mentioned that connecting the syllables of surnames, forenames, and geographic names makes them stand out and more recognizable. Also, this practice is consistent with records already in the database, which would make conversion easier. Opposing opinions stated that keeping all syllables divided would simplify cataloging and searching.

(C) that, terms for jurisdictions and topographical features would be separated from geographic names

52 of 63 total responses (82%) supported this proposal, 5 (8%) opposed it, and 6 (10%) had no comment. Of the ten teaching faculty members, nine supported the proposal, and one had no comment.

Those supporting this proposal felt that it would avoid confusion. It is easy for users to know that “Shanghai” and “Shanghai shi” are the same place, while “Shanghaishi” could be considered to be a different place from “Shanghai” and cause authority difficulties. One of the opposing views was against connecting any syllables of surnames, forenames, and geographic names and maintained that all syllables should be divided. Another opposing opinion said, “users do not make distinctions between geographic names from terms for jurisdictions and topographical
features. . . Why make it complicated with one more rule?” Some also suggested adding a hyphen between geographic names and terms for jurisdictions and topographical features.

(D) that, where the preferred form in LC Name Authority file (LCNA) is a Wade-Giles form, the preferred form in the pinyin database should be that Wade-Giles form converted to pinyin.

Of a total of 63 responses, 50 (80%) supported this proposal, 7 (11%) opposed it, and 6 (9%) had no comment. Of the ten teaching faculty members, nine supported the proposal and one had no comment.

Consistency was the main consideration for those who support this proposal. Some supporters also suggested that following this practice would make the conversion from Wade-Giles to pinyin easier.

2. When the Library of Congress applies the proposed (pinyin) word division,

(A) what effect do you feel the proposed approach to word division might have on your daily work, or on your institution’s operations, or on the public you serve?

Some representative responses are

(1) Wait and see, but whatever is done must be acceptable with current library operations.

(2) The LC proposed word division is still not obvious, nor standard, but it would be easy to teach.

(3) The proposed approach to word division does not have too much effect on my daily work; it’s just a matter of getting used to it.

(4) Cataloging would be easier, would conform to internal institutional practice, and simplify searching for the public served by bringing it more in line with how Western languages are searched.

(B) what advantages or disadvantages do you perceive from the standpoint of acquisitions, cataloging, searching, or meeting the needs of the users, etc.?  

Some representative responses are

(1) We would be moving toward international uniformity. Standardization on an international level can only make access easier.

(2) Pinyin romanization is shorter than Wade-Giles romanization, so in that aspect, pinyin has the advantage over Wade-Giles.

(3) The LC proposed pinyin word division is easier for most users.
(4) Using pinyin would be in line with what vendors currently use in acquisitions.

Other Word Division Suggestions and Issues

The questionnaire included a few other questions related to word divisions. Most responses did not give specific reasons for supporting or opposing.

1. Syllables should always be divided from each other with NO exceptions.

Of 63 responses, 15 (24%) supported this suggestion, while 41 (65%) opposed and 7 (11%) have no comment. Of the ten teaching faculty members, two supported the suggestion; 8 opposed it.

The main reason given in support was simplicity.

2. Should hyphens be used for word division?

Of 63 responses, 9 (14%) said “Yes,” while 51 (81%) said “No” and 3 (5%) had no comment. Two of the 9 “Yes” responses were conditional—they felt that hyphens should be used only when combining geographic names with jurisdictions and topographical features. One response opposing this suggestion stated that it was “too weird, and different from our current practice.” One of the ten teaching faculty members responded “Yes” to the question; 9 said “No”.

3. In dealing with the two sounds requiring the umlaut (i.e. lù and nù)

A total of 69 responses (two teaching faculty members each gave three responses and two library professionals each gave two responses) were distributed as follows:

Sixteen (25%), including three of the faculty members, suggested ignoring the use of the umlaut, giving as a reason that “in English language, different tones or sounds do not affect filing, so long as the spellings are exactly the same.”

Thirty-seven (60%), including seven of the faculty members, wanted to continue to use the umlaut. The main reason is that “it’s an important distinction. It’s part of the pinyin standard.”

Seven (11%), including two faculty members, suggested using a double “u” in place of the umlaut.

Two (3%), all faculty members, suggested adding “*e” to “u” in place of the umlaut.

Three (5%) had other suggestions: “lü-->lyu, nü-->nyu,” “Whatever practice is adopted, it should be used universally;” “or, follow the Chinese practice.”

Three (5%) had no comment.

4. An apostrophe should be placed between syllables when ambiguity exists as to whether a
middle consonant goes with the first or second syllable.

48 of 63 responses (76%) supported this suggestion, while 14 (22%) oppose it and 1 (2%) had no comment. Of the ten teaching faculty members, eight supported the suggestion and two opposed.

Reasons for supporting this suggestion were: fewer ambiguities; it is part of the pinyin standard, and so should not be changed. An opposition response mentioned that if proper names are regularly joined, there wouldn’t be much possibility for confusion with another place or person.

5. Tones should be indicated to distinguish meaning

Of 63 responses, 15 (24%) supported this suggestion, while 47 (74%) opposed it, and 1 (2%) had no comment. This issue was another of significant difference between library professionals and teaching faculty members. Of the fifteen responses supporting this suggestion, seven were from faculty members (=70% of faculty members). 83 per cent of library professionals were against this suggestion, while only 30% of faculty members opposed it.

A major reason given in opposition is that indicating tones might mess up indexing and confuse patrons. It would be difficult for catalogers to apply in a standard way.

Further Comments or Suggestions

Most responses to the survey offered further comments and suggestions. Those not mentioned in earlier paragraphs follow.

1. The conversion from Wade-Giles to pinyin romanization is long overdue. Let’s get on with the job and begin conversion as soon as possible.

2. It will be extremely important in deciding whether or not to make this change: (1) to hear what the Library of Congress and the networks plan to do about converting existing records; (2) to discuss how and at what cost we will be able to convert local online catalogs; and (3) to hear potential solutions to classification problems raised by the change.

3. There is no choice at all to change general and official rules of pinyin, such as hyphens, umlauts, tone numbers, etc. The official practice of pinyin is quite clear, and the Library of Congress should not mess with it.

4. Ideally, we should use aggregated pinyin (without hyphens) as shown in standard PRC dictionaries. But there are areas where the practice is unclear, therefore, “I support the LC proposal as a compromise which also has the advantage of being veyr close to our current practice.”

5. I hope that the Library of Congress can use a kind of software program to convert existing records to pinyin as Karl Lo has done for OCLC. “We will start to do new records.”
Statistical Report of the
Survey on *Pinyin* Romanization

**80 copies of questionnaire mailed to North American libraries on 1/13/97
**67 responses received by 2/20/97 (65 via survey questionnaire and 2 via email)

This survey is completed to represent the views of:
(17) an institution (16 questionnaires + 1 email)
(50) an individual (49 questionnaires + 1 email)
**including 10 questionnaires from professors

Feasibility of conversion

1. Do you agree that the Wade-Giles romanization should be converted to *pinyin*?
   (56) Yes  (6) No  (5) No comment
   84%—  9%—  7%+

   **including 10 “yes” answered by all 10 professors
   **4 out of the 6 who answered “no” did not further complete the questionnaire

LC Word Division Proposal

1. Do you concur with these aspects of the Library of Congress proposal?

   (A) that generally, syllables would be divided from each other.
   (52) Yes  (10) No  (1) No comment
   82%+  16%—  2%—

   **including 3 “yes” and 7 “no” answered by professors
   **answers from librarians and professors quite different (70% of professors said “no” while over 92% of librarians said “yes”)

   (B) that, only the syllables of surnames and forenames, as well as geographic names, would be connected. For example: Zhang Daqian, Deng Xiaoping, Ouyang Xiu, Beijing, Guangzhou.
   (52) Yes  (10) No  (1) No comment
   82%+  16%—  2%—

   **including 2 “yes,” 7 “no,” and 1 “no comment” from professors
   **answers from librarians and professors also quite different (70% of professors said “no” while over 94% of librarians said “yes”)

   (C) that, terms for jurisdictions and topographical features would be separated from geographic names. For example: Hangzhou shi, Wuxi xian, Changjiang, Emei shan.
(52) Yes  (10) No  (1) No comment
82%+  16%-  2%
**including 9 “yes” and 1 “no comment” from professors

(D) that, where the preferred form in LC Name Authority file (LCNA) is a Wade-Giles form, the preferred form in the pinyin database should be that Wade-Giles form converted to pinyin.

(50) Yes  (7) No  (6) No comment
80%-  11%+  9%
**including 9 “yes” and 1 “no comment” from professors

Other Word Division suggestions and Issues

1. Syllables should always be divided from each other with NO exceptions.
   (15) Agree  (41) Disagree  (7) No comment
   24%-  65%+  11%+
   **including 2 “yes” and 8 “no” from professors

2. Should hyphens be used for word division?
   (9) Yes  (51) No  (3) No comment
   14%+  81%--  5%--
   **including 1 “yes” and 9 “no” from professors
   **2 out of the 9 “yes” are conditional (i.e. only used in these cases: Hangzhou-shi, Emei-shan, etc.)

3. In dealing with the two sounds requiring an umlaut (i.e. lű and nű):
   **2 professors each gave 3 answers and another 2 librarians each gave 2 answers

   (16) Ignore the use of umlaut
   25%+  **including 3 from professors

   (38) Continue the use of umlaut
   60%-  **including 7 from professors

   (7) Use double u in place of umlaut (i.e. luu and nuu)
   11%  **including 2 from professors

   (2) Add *e to u (i.e. lu*e and nu*e)
   3%+  **all from professors

   (3) Other
   5%-  **lű→lyu, nű→nu
   **Whatever practice is adopted, it should be used universally
   **or, follow the Chinese practice
4. An apostrophe should be placed between syllables when ambiguity exists as to whether a middle consonant goes with the first or second syllable (e.g. Yan’an).

(48) Agree (14) Disagree (1) No comment
76%+ 22%+ 2%-

**including 8 “agree” and 2 “disagree” from professors

5. Tones should be indicated to distinguish meaning (e.g. wu1, wu2, wu3, wu4).

(15) Agree (47) Disagree (1) No comment
24%+- 74%+ 2%-

**including 7 “agree” and 3 “disagree” from professors

**answers from professors and librarians quite different (70% of professors “agree” while more than 83% of librarians “disagree”)

Prepared by Amy Ching-Fen Tsiang
APPENDIX

Pinyin Romanization: Word Division Recommendation

Library of Congress recommends following the National Library of Australia practice.

New developments in technology and computer programming have prompted the Library of Congress to again consider converting from Wade-Giles romanization of Chinese to the pinyin system. The Library has begun discussing options for conversion with the major utilities and the library community.

In 1990, the Library of Congress indicated that it wished to explore, with OCLC and RLG, possibilities of machine conversion of existing MARC records to pinyin. It was recognized then that a preliminary step in preparing for conversion would involve discussion and agreement on a standard for word division.

It has been almost forty years since the People’s Republic of China adopted pinyin. Pinyin is now generally recognized as the standard for romanization of Chinese throughout most of the world. And yet, there seems to be no generally accepted international standard for pinyin word division. Although the government of China has issued standards for word division, publishers and authors often do not conform to its guidelines. Dictionaries published in China do not follow consistent word division practices.

European romanization practices and word division practices also seem to vary greatly. For example,

- The British Library appears to romanize in pinyin, separating all syllables from each other with spaces.

- The University of London School of Oriental and African Studies uses Wade-Giles.

- The East Asian Department of the Staatsbibliothek zu Berlin uses pinyin, aggregating syllables, and then separating the resulting words from each other with hyphens.

- The Staatsbibliothek Preussischer Kulturbesitz Ostasienabteilung’s Far Eastern catalog uses Wade-Giles, but separates the individual syllables from each other with hyphens.

- The Bayerische Staatsbibliothek (Munich) catalog of the Far Eastern collection uses Wade-Giles as we do.

We feel that the absence of an international standard for word division provides us with the opportunity to propose a system that best meets our needs.

The National Library of Australia (NLA) has sought to maintain consistency in applying syllable aggregation practice as it converts its files from Wade-Giles to pinyin. NLA pinyin word
division guidelines are embodied in two of its CJK (Chinese, Japanese, Korean) Technical Committee resolutions, and read as follows:

That, where a cataloguer inputs Pinyin data into the National CJK System, each Chinese character should be input as one Pinyin syllable, except for proper and geographic names, where the syllables should be joined.

That . . . In the Pinyin database:

- where the preferred form in LCNA is a pure Wade-Giles form, the preferred form in the Pinyin database should be that Wade-Giles form converted to Pinyin

- where the preferred form in LCNA is an “established” form i.e. not a pure Wade-Giles form, for instance Chiang Kai-shek and Confucius, the preferred form in the Pinyin database should be that “established” form.

Note: LCNA is the Library of Congress Name Authority file.

In other words, individual syllables of surnames and forenames, instead of being hyphenated, would be connected, as would individual syllables of geographic names. Terms for jurisdictions and topographical features would be separated from geographic names. All other syllables would be separated from each other.

The Library of Congress proposes following the practice that has been adopted by the National Library of Australia. We believe that this approach would offer a number of distinct advantages:

1. It is familiar. It offers the least possible change from our present practice, and therefore would be easy to learn and apply.

2. It is straightforward. It eliminates the use of the hyphen entirely. The guidelines for its application can be written in a few words, making it easy to communicate and put into practice.

3. As NLA converts its database from Wade-Giles to pinyin, it is also converting hundreds of thousands of American MARC records. By adopting the same system of word division, libraries in this country, and major utilities, could conveniently utilize Library of Congress bibliographic records just as they have been converted by NLA, without having to take the further step of accommodating different word division practices.

4. Separating most syllables would make it possible for records to be changed to suit other institutional needs (i.e., it would allow for future syllable aggregation or connection).
5. It does not seem to present a conflict with an established Chinese or European policy or practice (because practices vary so greatly).

For all the above reasons, we believe that this would be the most economical approach for the Library of Congress, the major utilities, and individual libraries. Furthermore, it would be easy for users to learn and apply.

This proposal has been submitted to professional organizations for review and comment. The Library has also solicited comments on this proposal from scholars and other library users.