

ISSUES AND PROSPECTS IN EAST ASIAN LIBRARIANSHIP

Susan MacDougall
Canberra, Australia

This paper is a broad overview. Its purpose is to identify and outline current issues and trends in East Asian librarianship, some of which will be treated by other speakers in greater depth. Together the papers should provide a springboard for discussion and an indication of future directions.

There are basically two types of issues. The first type is temporary issues which can be solved by technological developments: - telecommunications solutions in the form of greater bandwidth and more ports; hardware solutions in the form of faster processing, larger hard disk capacity and more memory; and software solutions, which depend on software developers being convinced of the need for particular software developments. Then there are more enduring issues. They relate to attitudes, politics and policies. They are enduring because people can be intransigent. They need to be convinced of the need for action. Of course, technological issues and their solutions are influenced by politics at any level, so the distinction between these two types of issues is not clear cut, and decisions may be made for political reasons which are not always the best solutions to the problems.

Many of the issues for librarianship as a whole apply equally to Asian and East Asian librarianship. At all levels, these are influenced by management attitudes and priorities. They are: funding, resource sharing, interlibrary loan/document delivery, and adapting to change (including the Internet and its impact on library services). The vexed issue of how far information and access to information should be freely available is also a topic of debate. More relevant to multilingual librarianship are: the availability of staff with the appropriate combination of skills; bibliographic control; access; and, national standards versus global standards for MARC format, character sets and romanisation. These issues are also often discussed in broader arenas, but are relevant to East Asian librarianship too. Not surprisingly, issues of the English-speaking world tend to dominate in English-language conferences. There appears to be a gap in practice between North American and Australian libraries on the one hand, and the rest of the world.

I will expand on some of these issues in turn in so far as they relate to East Asian librarianship, and comment on what is being done about them. Our geographic coverage is China, Japan, Korea and Mongolia. Mongolia has not featured much in the past because of low levels of collection in Australian libraries. It would appear that the Mongolian book trade is not particularly active. Certainly, books about the Mongolian language tend to be written in Russian, while Inner Mongolian serials tend to be published in Chinese with a transliterated Mongolian title. However, Mongolian-language materials do exist in Asian collections and should not be overlooked. There are several Mongol languages, the official language of the Republic of Mongolia being Khalkha Mongol. In Inner Mongolia, which is under Chinese rule, the most

common Mongol language is Oirat. Now that the former dependencies of the Soviet Union have become independent, we may see a trend towards the resumption of the Mongol writing system, which was based on the Arabic script but written downwards, from top to bottom, under Chinese influence. It is interesting in being the only script written downwards from left to right. We could even see a push for the automation of the Mongolian script.

The Internet was mentioned as an issue. East Asian librarianship in Australia benefits greatly from the communication channels opened up by the Internet, because East Asian librarians are scattered and isolated. The Internet should improve timely access to online resources for information professionals and users alike in their search for information.

But not only must access be available: the documents must be readable, whatever the language or script. Over the Internet we can now look at library catalogues in Chinese, Japanese and Korean (CJK) scripts, and receive e-mail in Chinese, Japanese or Korean, provided we have the necessary software to view them. The current problem with CJK materials is finding software that will handle all three, rather than just one at a time. Such software certainly exists for some functions, although possibly not yet for e-mail. The diversity of non-compatible character sets and input methods is a problem. They tend to be developed separately by countries who may not be prepared to talk to each other. This, in turn, is a reflection of what happens in the computer industry, where companies are in competition rather than pulling together.

While the Internet opens up channels of access to end users, it offers poor subject indexing and poor search facilities in comparison to the powerful command languages of the traditional, pre-Internet search services. The result is high retrieval, low precision, and sometimes a complete failure to find relevant material. How much more difficult retrieval must be when there is a variety of languages, scripts and romanisations.

Nevertheless, there are major library networks offering multilingual capabilities. Cooperative networks such as the Research Libraries Information Network (RLIN) and the Online Computer Library Center (OCLC) already support CJK and some other non-Roman scripts, and have done so for several years.

There have also been some recent initiatives in Australia. The National CJK Service offers parallel bibliographic databases, one in Wade-Giles and one in Pinyin, both containing CJK scripts. The parallel databases involve duplication, not only of Chinese records but also of Japanese, Korean, and other mixed-language records. New records are converted regularly between the two, so that the databases mirror each other. The justification for this approach is the split in the library and user community between the use of Wade Giles and Pinyin romanisation systems. Both romanisations are in use in Australian libraries, and there is a demand for cataloguing copy in both. The National CJK Service is a bold and practical answer to a pressing problem here and now. In future, if the Library of Congress adopts Pinyin as a standard, the Pinyin version of a quarter of a million CJK bibliographic records is already available.

The dual system is at a half-way stage, foreshadowing the eventual changeover to Pinyin romanisation system for bibliographic records. The excellent work of C.P. Tang at the National Library of Australia as database manager should be acknowledged.

There have been other initiatives in Australia, in the form of three national roundtables on Asia and a Korean Working Party. The latest roundtable resulted in recommendations for action to redress problems relating to bibliographic control and access to Asian-language materials, collection development, and non-Roman scripts. The regional focus of the Third National Roundtable was on South Asia. Other issues which arose were: education and training for Asian librarianship and the preservation of Asian materials. The Korean Working Party had a brief to define future collecting directions and responsibilities among Australian libraries with Korean collections, and these focused on cooperation and resource sharing. Meetings resulted in a laudable cooperative agreement between the National Library of Australia and the Australian National University Library. Generally speaking, in a time of restricted funding, resource sharing becomes more necessary and, paradoxically, harder to achieve, since every library must put its primary users first.

Are national standards compatible with a global approach? There was a time when each library had its own individual practices. Automation and networking has brought standardisation within a country. Record sharing between countries brings further standardisation on an international level, as we see with the discontinuance of AUSMARC and the wholesale adoption of USMARC in Australia. There are now two major MARC standards - USMARC and UNIMARC. China is using UNIMARC for Chinese-language materials and USMARC for Western-language materials. This is the prime example of the split between the two. How can they be integrated into a single database?

This leads onto the question of multilingual automation, which is still an issue, especially in the integration of CJK scripts into mainstream databases. The sheer number of writing systems in the world and the size of character sets, particularly non-phonetic sets, creates storage problems, especially as the storage addresses of national character sets may conflict with each other. The problem is mainly political and relates to the autonomy and sovereignty of nations. It can be solved by technology and software with an international approach.

Character sets are one aspect of multilingual automation, and one of the stumbling blocks. Unicode, which was heralded as being the answer, is now considered to need enhancing from 2 byte codes to 3 byte codes. Professor CC Hsieh, at the 1996 Association of American Libraries (AAS) Conference in Hawaii, urged the International standards Organization (ISO) to adopt his proposed input methodology for creating those characters which were not present in the Unicode set. Professor Hsieh's proposed solution was based on the use of Standard General Markup Language (SGML), and is one that might be expected to end Unicode's ongoing problems in obtaining support from potential sophisticated users.

Allied with the automation of the world's scripts is the question of romanisation. Romanisation is seen as an easy alternative to displaying vernacular scripts, but it is not without its own problems, since several Library of Congress romanisation standards are not those adopted in the mother countries. This is not surprising, since the original purpose of romanisation was to indicate pronunciation of a language to non-speakers using their spelling conventions. The role of romanisation has changed. In the online environment it is now not so much a pronunciation guide as a general substitute for the original script and a search tool. In library circles in some English-speaking countries, the Wade-Giles versus Pinyin debate still continues. Pinyin will continue to be a second official Chinese romanisation for a while, and eventually replace Wade-Giles. For Japanese, the Modified Hepburn romanisation is different from the Japanese romanisation used by the National Diet Library of Japan. Other romanisation systems might also possibly become contentious, such as Greek romanisation.

But it is not just a matter of whether to romanise, but how far to romanise. USMARC has the entire record romanised with the addition of vernacular fields. In the era of the card catalogue, only the access points were romanised.

While it is a nuisance, conversion between romanisation and transcription schemes is not unsurmountable., There are now OPACs containing Chinese characters and Wade-Giles romanisation which can be searched via Wade-Giles OR Pinyin (for instance, at the University of California at San Diego). To do this, a programmed filter is used that converts the Pinyin search string into Wade-Giles. Karl Lo is active in this area.

Automatic conversion between Wade-Giles and Pinyin is not in itself a real problem, but more human intervention and editing is required where multilingual bibliographic records are involved. A typical conversion program uses a table of syllables which it recognises as uniquely Wade-Giles, but these syllable may also be words in other languages. It either converts them automatically or posts them as "doubtfuls" which go to human review. Examples which spring to mind are 'men' (Chinese/English) and 'to' (Chinese/English/Japanese).

Why use romanisation at all in records? Romanisation has been a pragmatic solution to the difficulties of handling data in other scripts, especially in an automated system. It has also provided access points which are interfilable in the local catalogue. On a global level, it seem rather parochial and inward looking to romanise another country's writing system. How would you like to find English titles transliterated into Cyrillic in a Russian catalogue? While romanisation has its uses, it is unacceptable if it makes documents unintelligible to native and non-native speakers alike. It also brings with it a new set of problems if the original script does not have spaces between words or upper case.

Along with romanisation, word division is a problem for several languages, including Japanese and Chinese (and Thai, for that matter). The solution for

filing Japanese in the card catalogue used to be a letter-by-letter approach, ignoring spaces. In the online catalogue, mixed filing sequences are unlikely to be used. They would certainly be confusing for the searcher.

Word division is prescribed for Wade-Giles romanisation, but is less standardised for Pinyin. Various suggestions have been made for combining, or "aggregating", Pinyin syllables with such symbols as hyphens or 'plus' signs. The National CJK Service has adopted the approach of maintaining the same word division as for Wade-Giles, removing the hyphen in order to form compound syllables for geographic and personal names. This does not necessarily conform to other practices, such as PRC Standard GB 3259-92 for transliteration and word division. However, it is a pragmatic approach to aid the conversion process.

And in fact, it does not really matter, according to Karl Lo, who explains that in most cases separate syllables can be combined and re-divided into single syllables with appropriate software. For that reason, he suggests, it is unnecessary to use hyphens or plus signs to combine syllables. For example, even a long name such as "Zhong guo ke xue yuan wu li yan jiu suo" may be combined without hyphens in any manner and can still be systematically and accurately separated into single syllables. In cases of ambiguity, apostrophes, rather than hyphens, should be used as division marks. Karl's examples are 'piao' or pi'ao'; 'xian' or 'xi'an'; 'chang'an' or 'chan'gan'. There may be an issue of the ease of reading. Whether 'Hanyu pinyin fang'an' is easier to read and understand than "Han-yu pin-yin fang'an" or "Hanyupinyin fang'an" is probably a matter of personal choice.

Karl is probably over-optimistic in expecting software to solve all problems, given the complexities and ambiguities of language. From personal observations of the conversion and reviewing process at the National CJK Service, at least 10 percent of converted records are sent for human review because of unresolved ambiguities. Matters are relatively simple in a Chinese-only environment, but complicated in a large, multilingual database. Moreover, the quality of the data depends very much on the cataloguer putting correctly-spelled data in the correct fields, subfields, with the correct diacritics and applying syllable separators appropriately.

I should like to add a comment here that tone markers make Pinyin unreadable. They are nearly as bad as diacritics, which Pinyin is rescuing us from. I am unsure whether they are covered in the standards, and I understand how they can contribute to differentiation between homophones in a tonal language. However, I really think that the inclusion of Chinese characters is the only ultimate way to make text really understandable.

Standards vary for descriptive cataloguing, access points, capitalisation and punctuation. The artificial punctuation between fields and subfields may not be acceptable in non-Western countries. The concept of author as main entry may not be acceptable either. For these reasons, practices vary from country to country, and some sort of common interface or conversion is required to transfer data from one country to another.

The Australian national roundtables on Asia identified the lack of staff with combined library, language and computing skills, resulting in a lack of bibliographic control of some Asian language materials. Libraries with small Asian collections and a small, all-purpose, staff find cataloguing these materials a particularly difficult problem.

One solution is specialised services with concentrations of expertise offering multilingual cataloguing. The Cooperative Action by Victorian Academic Libraries (CAVAL) is already heavily involved in cataloguing non-English language materials in some forty different languages. They are now surveying the demand for a CJK cataloguing service. This is another instance of the already existing trend towards the "outsourcing" of cataloguing.

Looking ahead, globalisation will continue, and it will bring greater standardisation. However, since each country must have autonomy over its own processes and procedures and put its local users and needs first, development will be on a national basis. Cooperation must be on a voluntary basis of enlightened self-interest or mutual advantage. Of course, there are broader alliances such as the European Community which are a unifying factor. Catalogue record sharing and access to international resources go hand in hand with international standardisation. Yet how can you download and integrate records in different formats, some all in one script, some in a mixture? Thus, while there will be a trend towards standardisation, diversity will also have to be supported by means of umbrella software with common interfaces and conversion software. As long as the data within a system is consistent and accurate, it should be possible to convert it automatically with some human intervention. Therefore, accurate data entry and quality of data becomes extremely important.

Do East Asian library associations have a role to play in future developments? Would a group detached from individual libraries be in a better position to identify and lobby for improvements in East Asian librarianship? Do small groups have enough influence to be effective? Is there a role in identifying and indexing worthwhile Internet resources - a role of sifting and quality control? In the United States, Ken Klein certainly sees a role for the Council on East Asian Libraries (CEAL) to assume greater responsibility for systematic control over the East Asian related materials available on the World Wide Web by critically evaluating and authoritatively indexing those resources. This may be something that EALRGA should be considering too.

The foregoing are some of the issues which currently confront us. For the most part, they involve standards and transborder communication in an increasingly international environment. The challenge is how to bring diverse systems and standards together. It is possible to be spectacularly wrong when making predictions, since it is difficult to anticipate triggering technologies which unexpectedly make new advances possible. Although most of the issues discussed involve technological advances, the fact remains that only with the support of library managers can forward planning and resourcing problems be solved.