

# **Information retrieval of Korean materials using the CJK bibliographic system: issues and problems**

Jung-ran Park

*Indiana State University, University of Hawaii at Manoa*

The acceleration of both print and electronic information production and use, along with the expansion of information utilities such as the on-line databases and the explosive growth of the Internet, are all clear indications of society's inevitable march toward the knowledge-based information age. The fundamental issue of information retrieval in this information age is closely interrelated with the characteristics of language.

In this paper, I will focus specifically on the Korean component of the CJK (Chinese, Japanese and Korean) system, one of the electronic bibliographic utilities. From a background on Korean linguistics, I will examine the extent to which this system accurately reflects the informational needs of its users from the perspective of information retrieval and access. In particular, I will focus on the inherent drawbacks of the McCune-Reischauer (MR) romanization scheme, as well as word division practices, which impede the full exploitation and use of the informational tools of the CJK system.

The phonetic-based MR system causes inconsistency, ambiguity and idiosyncrasies in the searching for and retrieving of information by users, as well as in creating/cataloging information by catalogers. These problems deteriorate both information retrieval and the process of creating new tools, such as indexes and thesauri. The result is that the goal of bibliographic control is defeated. This brings about various hindrances for Koreans in accessing Korean language materials in academic and public libraries in North America.

The need for indexes and thesauri will be unavoidable in the future. For these purposes, a romanization scheme based on morphemic principles has much merit, in that such a system dissolves semantic ambiguity, while the MR system increases such ambiguity. Also, a system requiring fewer diacritics would be much more suitable to the electronic environment. Concerning word division, an automatic parser needs to be implemented to ensure critical factors, such as consistency, quality, and efficiency, in information retrieval. The development of an automatic parser needs to be based on linguistic principles, the foundation of natural language processing.

## **1. Introduction**

The increase in both print and electronic information resources, the expansion of information utilities such as the on-line databases, and the explosive growth of the Internet, are all clear indices of society's inexorable march toward the knowledge-based information age. As a social and societal institution, the library is naturally reflective of this fundamental societal shift.

Historically, the principal role of the library, among others, has been as a conservatory of the human record. However, the fundamental societal shift mentioned above has forced libraries to rethink and redefine their role. Along with its traditional role, the library as an institution has taken on other functions as a result of the information revolution underway. Among these are: providing access to and dissemination of electronic information resources, fulfilling users' information needs, and empowering information seekers by providing instruction and methods for the accessing of new electronic resources. One of the great figures in the area of Library and Information Sciences, Jesse Shera, back in 1972 foresaw this role of libraries thus: "maximizing the graphic records of the social utility". After all, the function of information access and information dissemination mirrors the needs and demands of our society, which is increasingly knowledge and information based.

Information access and dissemination is closely tied to the needs of information users. Therefore user needs must be ascertained through study and analysis of their information-seeking patterns and behavior. In this regard, user studies provide an important avenue for improving existing information systems and for designing new systems and components, such as electronic databases.

According to Bonta (1992), among international students in North American universities and colleges, more than half are East Asian, mainly from Korea, China (and Taiwan), and Japan. The size of this East Asian segment of the student body will inevitably grow in the future. Among the factors contributing to this trend are the strong long-term economic growth potential in these countries (despite short-term reversals), the Confucian-based emphasis on education, and the well-established and continued perceived prestige and value of an overseas degree, especially on the graduate level. With the recognition of the critical nature of these countries' resources, and the growing informational needs in this area, the CJK (Chinese, Japanese, Korean) system was developed by the on-line utilities, RLIN (Research Library Information Network) and OCLC (Online Computer Library Center).

In this paper, I will focus specifically on the Korean component of the CJK system. In the following sections, I will examine the extent to which this system accurately reflects the needs of its users from the perspective of information retrieval and access. In particular, I will focus on the types of drawbacks which hinder the full exploitation and use of the informational tools of this system. These drawbacks block not only the use of information, but impede the sharing and cooperation necessary for the building of bibliographic databases. After consideration of these critical impediments, I will suggest some potential remedies that could serve to overcome these problems. In sum, the goal of this paper is to point to the need for a serious reexamination of the current system.

## **2. Drawbacks of the CJK Korean system**

The chief, although by no means exclusive, users of Korean materials in North American libraries are native Koreans. Let us assume that the Korean collections at, for example, Hamilton Library at the University of Hawaii or at McCully Public Library in Honolulu primarily serve patrons who can at least read Korean. This assumption could be validated by surveying the use and perceived utility of these collections among non-native Korean speakers who can read Korean, non-native speakers who cannot read Korean, and native Korean speakers themselves. My submission is that such a survey would show that the drawbacks of the Korean cataloging system are primarily manifested in its deficient utility for the primary,

namely Korean, users of the CJK bibliographic utilities. The reasons for this will become clear below.

The prevailing situation is a result of the fact that the Korean cataloging system, based on the McCune-Reischauer (MR) romanization scheme, is designed for and aimed at non-native speakers, especially English speakers. The MR system, although based on the phonetic structures of Korean, is designed to accommodate non-native Korean speakers. Phonetic types of romanization, such as the MR scheme, represent speech sounds as they are pronounced. This brings about multiple hindrances for Koreans in the accessing of Korean materials, as I will describe in the following sections.

This situation also causes impediments in the development of tools for the enhanced retrieval of information. It is to be hoped that the acceleration of technological development will extend into the area of materials for Korean and, for that matter, other non-Roman scripts. The need for more efficient information access and retrieval through the use of such tools will only increase as a result of the information explosion. From such a perspective, the current system is clearly not adequate.

As well, the current system is neither cost-effective nor time-effective. It blocks the sharing of bibliographic records and the cooperation necessary for the efficient building of databases. In other words, using the current system Korean catalogers have to spend time creating new records rather than duplicating existing records, as the current system blocks even an adequate searching of records.

In the following sections, I will expand on the above indicated drawbacks, focusing on the areas of romanization and word division. However, it should be kept in mind that these two major problem areas are not meant to represent the entirety of impediments engendered by the current system. Within the scope of this paper, I will not examine the entire gamut of problems. However, other difficulties will be enumerated in the conclusion.

## **2. 1. The inherent problems of romanization**

At present, there is no unified romanization scheme in use in Korea. Among several systems, the McCune-Reischauer (MR) system, the Yale system, and the Ministry of Education scheme are employed in different sectors for different uses. For example, in the area of publication and bibliographic records in libraries, the MR system is used. Conversely, the Yale system is uniformly used by linguists within and without Korea. Lastly, government documents, including street signs and road maps, employ the Ministry of Education system.

The differences among the above cited schemes can be discerned in the linguistic principles which underlie them. The MR system is based on the phonetic structures of Korean, as the inventors clearly specified in their original presentation. Phonetic types of romanization represent speech sounds in the manner in which they are pronounced. For example, in English the word “two” is transcribed phonetically as follows: [tu]. This romanization clearly does not capture the semantic differences caused by the existence of homographic forms. Thus, if we romanize a language on the basis of its phonetic structure, we cannot differentiate among words sharing the same pronunciation, of which especially English, but also Korean, is burdened with. The following illustration in English makes this clear: “two”, “to”, “too” are all transcribed as [tu].

On the other hand, the Yale system is based on morphemic principles. The morphemic type of romanization transcribes the base form of a word regardless of

sound changes. Korean is one of the languages employing a rich morphophonemic complexity. In other words, the base form of a word changes according to the adjacent sound environment. Most agglutinative languages, including Japanese, fall into this category. They are all very complicated morphophonemically. For example, the form of the Korean word 'mul' (water) is changed into 'muri' when the nominative particle *-i* is attached to it. The morphemic type of romanization is not reflective of sound change as is the phonetic type of romanization, instead reflecting the base form. (The romanization system of the Ministry of Education is structured, if we can envision a continuum, toward the phonetic type. Owing to the scope of this paper, I will not detail this system here).

As mentioned, the current cataloging system dealing with Korean materials employs the MR romanization scheme. Noteworthy is the inventors' originally stated goal for their romanization scheme:

We have devised our romanization with the purpose of providing a comprehensible guide to the standard modern pronunciation of Korean for those unfamiliar with the language, as well as for those who know it. (McCune and Reischauer: 1940).

In the above passage, we can perceive that this system was primarily designed for users who cannot read Korean. The phonetic principle is very much valid for such a purpose. For those unfamiliar with Korean, such as scholars with a need for Korean materials, journalists, tourists, etc., this system could adequately, if not optimally, play a role.

However, the primary users of Korean collections in North America are those who are familiar with Korean, either native speakers or non-natives with at least a modicum of fluency, not those unacquainted with Korean. Non-native speakers using Korean materials can presumably be able to at least read basic vocabulary. The fundamental reason for the drawbacks of the Korean cataloging system lies in this set of circumstances.

In other words, because the MR scheme romanizes on the basis of the phonetic side of the non-native speaker's pronunciation, Korean users bear the resulting confusion. For example, the distinction between the voiceless [p] and voiced [b] is not heard by Koreans, while these are clearly distinct phonemes to English speakers. The MR romanization system follows this phonetic distinction.

Another problem engendered by the use of the MR system is that it causes information loss. This is especially true in the area of name authority. By romanizing according to pronunciation, a given name can become unclear. For example, the Korean name *Kim Sok-min* becomes *Kim Song-min* according to the MR system. With author names romanized according to the MR system, ambiguity becomes inevitable. Through informal experiments, I have noted that Korean students rely heavily on the author's name search in the retrieval of desired information, as search by title results in a lower rate of success, largely because of the MR system. However, Koreans discover that searching by the author's name inevitably involves this inherent ambiguity.

This could be very critical when considering the design of electronic indexes, especially name indexes. The linguist S. Robert Ramsey (1982) presented an insightful suggestion in this case: "This information loss becomes especially critical when all cataloging work is done by computer, and so it is perhaps time to give some

thought as to how appropriate McCune-Reischauer is in cases where precise data processing is required.”

Along this line, the MR system increases the number of homographs, the primary cause of semantic ambiguity. This phenomenon can be illustrated by the previously cited English example *two, to, too*. If these three lexical items are encoded according to their common phonetic form, *tu*, the resulting semantic ambiguity can be clearly seen. This is the case in many instances with the MR scheme. Considering the ongoing creation of thesauri for Korean databases, such ambiguity inevitably cause significant impediments in the process of information retrieval.

Lastly, the MR system brings about inconsistency in the creation of bibliographic records. In other words, when catalogers encode words according to their pronunciation, there is a strong possibility for individual catalogers to create inconsistent and arbitrary records. This is based on the fact that the pronunciation of a phrase can be represented differently according to speech register. If a cataloger pronounces a phrase using careful speech, the resulting transcription would be different from that of a transcription based on casual speech. The creation of differing records is thus entirely possible, either by the same cataloger, or different catalogers transcribing identical material. Furthermore, pronunciation can vary among catalogers based on dialect, even though the MR scheme has suggested rules to follow.

The following bibliographic record illustrates this problem. Owing to the scope of this paper, I will not indicate each field and subfield designation, or the use of diacritics.

	100 1	Kim, Yong-un,\$d1927-
	245 10	Che-2 <b>k<sup>^</sup>onggun<sup>gn</sup>on</b> :\$bkungmin kukka <sup>^</sup> ui wans <sup>^</sup> ong <sup>^</sup> ul wihay <sup>^</sup> o
/\$c		Kim Yong-un.
	246 3	Ch''op''an.
	260	S <sup>^</sup> oul T <sup>^</sup> ukpy <sup>^</sup> olsi :\$bChisik San <sup>^</sup> opsa,\$c1998.

### **Bibliographic record 1**

For the sake of illustration, let us take a look at the 245 field. This field represents the title of a bibliographic item. The portion of the title in bold, *k<sup>^</sup>onggun<sup>gn</sup>on*, represents the casual speech register. If the cataloger who created this record had pronounced it using careful speech, then it would have the same form as in *Han'guk*, in that the sound environment of the final underlined consonant of the first syllable, *kon<sub>u</sub>*, is the same as in *han<sub>u</sub>*. Both are followed by the velar sound [g]. In casual speech, however, the nasal sound [n] becomes assimilated into the following velar sound [ng]. In other words, in careful speech, the final consonant remains as a nasal sound, as indicated in bold: *k<sup>^</sup>onggun<sup>gn</sup>on* instead of *k<sup>^</sup>onggun<sup>n</sup>on*.

This inherent drawback hinders the ability of users to search the record. This hindrance applies to the professional cataloger as well, resulting in a major impediment to the sharing of cataloging records. Consequently, the goals of bibliographic control are vitiated, and the concomitant necessity for the creation of new records increases costs unnecessarily.

The MR romanizing scheme has inherent problems that are critical in the era of the information age. These problems deteriorate information retrieval as well as the process of creating new tools, such as indexes and thesauri. The minimum standards and criteria of bibliographic control cannot be maintained under the current system vis-à-vis Korean and similarly structured languages.

## 2. 2. Problems of word division

According to Yi Sung-u (1993), word division errors appeared in 29% of Korean standard books in current use in the school system. This highlights the difficulty in conducting word division in the written Korean form. I suspect that this problem can be found in any type of highly agglutinative language because of their inherent morphosyntactic complexity. Inconsistent and arbitrary word division can be found in even the most authoritative Korean dictionaries. This phenomenon is made worse in part by the frequent change of word division rules promulgated by the Ministry of Education in Korea. Interestingly, this problem has not attracted the attention of the general public. Only educators, publishers, linguists, and journalists are to some extent aware of and concerned with the issue. This is because arbitrary word division does not cause communicative hindrances in ordinary language use. That is, through context or conversation, any communicative ambiguities can be resolved.

However, the matter of word division is a critical factor in the electronic environment, where there is little or no context to resolve any semantic ambiguity. As well, at present there is no satisfactory software program to process the multiple properties of natural languages, such as accent, intonation, paraphrase, context, etc. In this regard, then, the fundamental problems of word division may exist in any type of romanization system.

According to the Library of Congress *ALA-LC Romanization Tables (1991)*, in addition to the rules specifying word division, there are four basic underlying principles. Before looking at some of the specific rules, I will go over one of the basic principles. The first basic principle is stated thus: “Each word or lexical unit (including particles) is to be separated from other words (ALA-LC ... 1991: p. 82)”. The following record helps us to understand this principle:

245 00	Y^oksa sok ^ui in'gan kwa chis^ong ^ul t''amgu handa /\$c Kim Chae-yong ...[et al.] p''y^on.
250	Che 1-p''an.
260	S^oul :\$bHan'gilsa,\$c1996.

### **Bibliographic record 2**

Referring to the title field (245), it can be seen that in the title *Yoksa^ sok^ ui^ in'gan^ kwa^ chisong^ ul^ t''amgu^ handa*, individual units are conventionally designated by the mark ^, denoting a total of eight spaces. This principle follows one of the suggestions presented at the 1981 Workshop Conference on Korean Romanization, held at the University of Hawaii under the auspices of the Korean Studies Center. The main aim of the conference was to examine the MR system with a view toward “ ... producing more uniform, more consistent, and simpler guidelines for those wishing to use this system to romanize Korean.” Under such a goal, a total of thirty-three recommendations were devised. Included among these is the basis for the unit division principle.

When looking at the totality of recommendations, we can see a conceptual basis for the MR system incorporating morphemic principles, in which romanization is performed according to the base form of a word. Let us look again at the ALA-LC principle: *Each word or lexical unit is to be separated from other words*. At the 1981 conference, a draft proposal by Dae. W. Chang of the Library of Congress (LC) was presented. Interestingly enough, his proposal reflects the desire of Korean catalogers to seek a system that “... more exactly represents the individual words of Korean so that scholars can work their way back to the original syllables ...” In other words,

Korean catalogers themselves are confused by the various and non-standardized ways used to romanize the same word, depending on the speech register used by the individual cataloger.

As the first step toward resolving this problem, the basic principle relating to word division has, in the span of eighteen years, been largely implemented. However, this principle engenders its own problem in that such word division is not used in Korean being contrary to conventional practices of the language. This could bring about a very heavy burden to Korean users in their searching processes. In fact, the above title has only three word divisions in the actual Korean written form, as follows: *Yoksasokui<sup>^</sup> in<sup>^</sup>gankwa<sup>^</sup> chisongul<sup>^</sup> t<sup>^</sup>amguhandu*. Moreover, this rule presents another intrinsic difficulty. It applies only to case particles of a noun phrase, not to affixes of verb phrases, as shown below:

245	10	Hakkyo ka ch <sup>^</sup> ongsony <sup>^</sup> on pihaeng e mich <sup>^</sup> in <sup>^</sup> un y <sup>^</sup> onghyang e kwanhan y <sup>^</sup> on <sup>^</sup> gu = \$bA study on school and juvenile delinquency /\$c[y <sup>^</sup> on <sup>^</sup> gu ch <sup>^</sup> aegimja Kim Chun-ho, y <sup>^</sup> on <sup>^</sup> guja No S <sup>^</sup> ong-ho].
246	31	Study on school and juvenile delinquency
260		S <sup>^</sup> oul :\$bHan <sup>^</sup> guk Hy <sup>^</sup> ongsa Ch <sup>^</sup> ongch <sup>^</sup> aek Y <sup>^</sup> on <sup>^</sup> guw <sup>^</sup> on,\$c1993.

### Bibliographic record 3

In the title field (245) of the record shown above, the part indicated in bold is an affix of the verb phrase *mich<sup>^</sup>i*. Thus, we can see here that the word division principle is not applied to the entire structure of the sentence. This could cause confusion for users in the process of information searching.

Concerning the specific rules for word division suggested by the Library of Congress, it can be said that contradiction rules. This seems to be a result of misunderstanding of the linguistic concepts underlying grammatical elements. I will not labor the matter here. However, It should be pointed out that such self-contradictory rules obviously engender inconsistent and arbitrary practices in the cataloging of Korean materials. Indeed, inconsistent practices and arbitrary cataloging records are the reality in the Korean collections of most North American libraries today.

Moreover, there are many Library of Congress cataloging records which embody flaws in word division. I will present several examples of such records below to illustrate this point. These Library of Congress records are structured contrary to the rules set by the Library of Congress itself for word division.

100	1	Kang, S <sup>^</sup> ok-ky <sup>^</sup> ong,\$d1951-
245	10	<b>Sup sok</b> <sup>^</sup> ui pang /\$c Kang S <sup>^</sup> ok-ky <sup>^</sup> ong.
250		Ch <sup>^</sup> op <sup>^</sup> an.
260		S <sup>^</sup> oul : <sup>^</sup> bMin <sup>^</sup> umsa,\$c1986.

### Bibliographic record 4

There are no rules specifying division of the word in bold *sup<sup>^</sup>sok* into *sup<sup>^</sup> sok*. In the Korean written form, there is no division between *sup* and *sok*. Thus, we can safely assume that this practice is arbitrary.

100	1	Ch <sup>^</sup> oe, S <sup>^</sup> og-y <sup>^</sup> ong.
245	10	Ilcheha <b>musongnon</b> kwa singminji kw <sup>^</sup> olly <sup>^</sup> ok /\$cCh <sup>^</sup> oe S <sup>^</sup> og-y <sup>^</sup> ong.
260		[Seoul] :\$bS <sup>^</sup> gy <sup>^</sup> ong Munhwasa,\$c1999.

### Bibliographic record 5

The above bibliographic record 5 is runs contrary to rule 5B: “If two single character substantives appear in succession, write the second one as a separate word (ALA-LC ... 1991: p. 85)”. As such, the word in bold in the 245 field should be separated in the following way: *musok<sup>^</sup> non*.

Let me gloss another bibliographic record as shown below:

100 1	Pak, Nam-ch <sup>^</sup> ol,\$d1953-
245 10	Pan sidaej <sup>^</sup> ok koch <sup>^</sup> al :\$bPak Nam-ch <sup>^</sup> ol sijip.
250	Ch <sup>^</sup> op <sup>^</sup> an.
260	S <sup>^</sup> oul-si :\$bHan Ky <sup>^</sup> ore,\$c1998.
300	141 p. ;\$c21 cm.
440 0	Han Ky <sup>^</sup> ore sinjak sijip

### Bibliographic record 6

The above example does not coincide with rule 3B: “Connect a prefix, such as the native Korean prefixes kat, hol, hat, p<sup>^</sup>ut, to the words that follow them (ALA-LC ... 1991: p. 84)”. The bold type in the 260 subfield (publisher information) and the 440 field (series) indicates the native Korean prefix *Han*. Even though it is not used as an example in the above cited rule, according to rule 3B it should be corrected to the following word, as *Han<sup>^</sup>gyore*. Interestingly, I could find a record in which there is no division between *Han* and *gy<sup>^</sup>ore*, observing rule 3B, as shown below:

100 1	Yi, Chin-ho.
245 10	Nongsakkun <sup>^</sup> ui my <sup>^</sup> ongham /\$cYi Chin-ho.
260	S <sup>^</sup> oul-si :\$Han <sup>^</sup> gy <sup>^</sup> ore Sinmunsa,\$c1994.
300	140 p. ;\$c21 cm.
440 0	Han <sup>^</sup> gy <sup>^</sup> ore sis <sup>^</sup> on.

### Bibliographic record 7

These examples illustrate the fact that inconsistent cataloging is being widely practiced when it comes to Korean materials, to the ultimate detriment of information users, primarily native Korean speakers. Due to inconsistent practices in word division, the recall percentage of retrieved information is inevitably lowered in the process of searching for Korean language materials. In addition, catalogers spend valuable time creating new records rather than exploiting the existing record, due to the inconsistent practices of word-division. Consequently, cost-effective cataloging practices are vitiated.

### 3. Addressing the drawbacks

Thus far, I have attempted to present the drawbacks inherent in the Korean cataloging system. I have addressed such hindrances both from the perspective of information retrieval by public users and from the perspective of professionals in the cataloging system. In order to suggest some alternatives, a survey and analysis of user information-seeking behavior is a definite prerequisite. In this regard, I am not in a position to suggest specific ways to resolve problems. Projects such as improved romanization schemes and word-division systems are enormous undertakings.

However, through observations and informal experiments, I have found that the current romanization scheme is definitely inadequate in fulfilling the needs of its users in that identical information can be rendered by differing romanization. To

address this matter, I strongly hold that a romanization scheme based on morphemic principles is necessary. In the previous section, I presented critical problems only from an information retrieval perspective. Thus, I did not illustrate other problems, such as the employment of diacritical symbols. Through surveying materials on Korean romanization, It is notable that the principal drawbacks of the MR system adduced by mainly linguists center on the matter of diacritics. One of the principal advantages of basing romanization on morphemic principles is that the need for diacritical symbols is substantially reduced.

As mentioned in the previous section, the need for indexes and thesauri will be unavoidable in the future. For these purposes, a romanization scheme based on morphemic principles has substantial merit in that such a system resolves most semantic ambiguity, while the MR system serves to increase the semantic ambiguity. Also, a system requiring fewer diacritics would be much more suitable to the electronic environment.

Concerning word division, I think that an automatic parser needs to be implemented in order to ensure critical standards, such as consistency, quality, and efficiency, in the process of information retrieval. As pointed out earlier, word division in Korean is very much flexible in actual written form. Thus, even though there are guidelines and rules for word division, inconsistent and arbitrary practices, with concomitant pitfalls, are the rule. The design of an algorithm for an automatic parser is beyond the scope of this paper. However, the development of an automatic parser needs to be based on linguistic principles, in that such principles form the foundation for natural language processing.

Lastly, I would like to point out the importance, indeed the necessity, for bibliographic instruction both for the native as well as for the non-native speaker. Bibliographic instruction is especially crucial to native Korean students in that they are not accustomed to reference and information services, library automation, and other prevailing standards and practices of North America libraries.

The hindrances inherent in the Korean cataloging system make this all the more critical. As sketched in the previous sections, there exist multiple impediments to adequate and efficient information retrieval. Besides the drawbacks forming the focus of this paper, there is another serious hindrance that should be briefly mentioned. This is the problem caused by translated subject headings. Translated subject headings inevitably hinder efficient information retrieval. The conceptual gaps that exist between any two languages, more so between languages as divergent as English and Korean, result in catalogued headings having some degree of inequivalent semantic value as a result of the translation process. This is an unavoidable reality; the goal should be to maximize awareness of this and other related phenomena in order to alleviate and possibly eliminate anticipated difficulties.

#### **4. Conclusion**

In many ways, the problems and hindrances outlined in this paper are not solely confined to the cataloging of Korean materials. In different ways and with different specifics, other languages employing non-roman scripts share similar impediments. Just within the CJK system, Japanese, which is grammatically and structurally close to Korean, has similar word division issues. Chinese, structurally distinct from both Japanese and Korean, is faced with a different set of issues but similar hindrances in information retrieval.

Bibliographic databases, search strategies, indeed the entire electronic information environment, were originally designed based on the Roman alphabetical

system. Therefore, using romanized Asian scripts in such an environment will inevitably cause some amount of difficulties. The goal should be to minimize these difficulties to the extent possible in order to maximize searching and cataloging efficiencies. Cost and time considerations are other factors that need to be optimized. My view is that the current system dealing with Korean materials does not adequately meet these criteria.

Specifically concerning Korean, a perfectly accurate representation of the language in romanized form is not attainable. It is fantasy to think that romanizing a given language precisely represents its phonemic value. When McCune and Reischauer devised their romanization scheme in 1939, the non-native phonetic values they assigned to it resulted in distortions and difficulties. Any such scheme aimed at the non-native speaker will never be adequate. Perhaps we should look to China as an example in this regard. Although circumstances differ, the Chinese themselves found the Wade-Giles scheme to be inadequate in representing their language. Therefore, Pinyin was promulgated as the superior alternative.

The problem of word division is potentially more easily dealt with. The rules set forth by the Library of Congress need to be modified to make their uniform implementation more easily attainable. To achieve this result, linguists and library professionals need to work in concert. However, the creation of an automatic parser is ultimately necessary to ensure consistent and cost effective cataloging, and to promote successful information retrieval.

Finally, I think it is imperative that students, faculty, researchers, and scholars be made aware of the issues and difficulties put forth in this paper. Most Korean foreign students enrolled in North American universities assume that the reason for inadequate retrieval when searching for Korean materials lies in poor search strategies on their part, lack of materials, and the like. The issues outlined in this paper are not generally known to them. The result can be detrimental long-term effects on their academic career. Bibliographic instruction for this group, and others who have the need to search for Korean materials, would go a long way toward improving scholarship and research in Korean materials. That should be the ultimate goal.

## References

1. Austerlitz, Robert et al. Report of the workshop conference on Korean romanization. In *Korea Journal*, 4: 111-125.
2. Bonta, Bruce D. 1992. American higher education and international programs: background, issues, and a future agenda. In *The role of the American academic library in international programs*, edited by Bruce D. Bonta, James G. Neal., Greenwich: Jai Press, p. 12.
3. Klein, Edward F. Romanization of Korean: do armchair linguists have the answer? In *Korea Journal*, 22: 20
4. Korean. In *ALA-LC Romanization Tables: Transliteration Schemes for Non-Roman Scripts*, approved by the Library of Congress and the American Library Association ; tables compiled and edited by Randall K. Barry. Washington: Cataloging Distribution Service.1991, Pp. 79-93.
5. Martin, Samuel E. 1968. Korean standardization: problems, observations, and suggestions. In *Ural-Altische Jahrbucher*, 40: 84-114.
6. McCune, G.M. and E.O. Reischauer. 1940. The romanization of the Korean language: based upon its phonetic structure. In *Transactions of the Korea Branch of the Royal Asiatic Society*, 29:1-55.
7. Mi, Sung-U. 1993. *Sae Machumpop kwa Kyojong ui Sirche*, Seoul: Omungak, p. 36.

8. Ramsey, S. Robert. Writing Korean with roman letters. *In Korean Journal*, 22: 33.
9. Shera, Jesse.1972. The role of the library in the dissemination of information. In *The foundation of education for librarianship*. Wiley, 163-194.
10. Shera, Jesse.1972. The role of the library in the social process. In *The foundation of education for librarianship*. Wiley, 135-162.
11. Sohn, Ho-min. 1986. Romanization of Korean: a cross-phonemic approach. In *Linguistic expeditions*. Seoul: Hansin Pub.
12. The Ministry of Education. 1979. The romanization of Korean.
13. Munhwa Kangwangbu kosi (The Ministry of Culture and Tourism of the Republic of Korea Announcement). 2000. Sae Romacha pyogipop (New System of Romanization for the Korean Language).