Cross-lingual Name and Subject Access

Mechanisms and Challenges

By Jung-ran Park

This paper considers issues surrounding name and subject access across languages and cultures, particularly mechanisms and knowledge organization tools (e.g., cataloging, metadata) for cross-lingual information access. The author examines current mechanisms for cross-lingual name and subject access and identifies major factors that hinder cross-lingual information access. The author provides examples from the Korean language that demonstrate the problems with cross-language name and subject access.

Today's global information society, benefiting from rapidly advancing communication technologies, spans geographical, lingual, and cultural boundaries. Recognition of the need for knowledge organization and integration, and access to cross-cultural and cross-lingual resources has greatly increased. The 2004 ISKO International Conference on "Knowledge Organization and the Global Information Society" and a 2004 special issue of *Cataloging and Classification Quarterly* ("Knowledge Organization and Classification in International Information Retrieval") are two examples.[1] International digitization projects have opened access to medieval texts as well as images and primary sources housed in libraries and institutions around the world, greatly advancing global access to multicultural resources.

The technological revolution that brought forth the global information society also has spurred recognition of the necessity for international collaboration aimed at multicultural education and diversity.[2] Linguistic and computational linguistic communities have collaborated in developing multilingual information resource discovery tools, such as concept-based indexing. These are used primarily for cross-lingual information processing. One example is *EuroWordNet,* which is based on Princeton University's WordNet, a lexical database for the English language.[3] The Open Language Archives Community (OLAC) has also been engaged in archiving, disseminating, and preserving language and cultural resources, including language-engineering tools, through utilization of the Dublin Core metadata standard.[4]

The challenges of accessing resources across cultures and languages suggest this is an area of particular interest to librarians, who are responsible for description and access. As a first step in exploring this topic, the author studied current practices in providing cross-cultural and cross-lingual information access. In this paper, she identifies problem areas and suggests directions for future study. The scope is limited to studies dealing with cataloging and metadata schemes for cross-cultural and cross-lingual information access.

Jung-ran Park (jung-ran.park@cis.drexel.edu) is Assistant Professor, The iSchool at Drexel, College of Information Science and Technology, Drexel University, Philadelphia.

The comments of the editor and anonymous reviewers regarding this paper greatly improved the quality of the work and I would like to express my appreciation to them. This paper is based upon a conference presentation, "Global Access to Cross-cultural and Cross-lingual Resources: Current Research Trends and Their Implications to LIS Curricula," at the 2007 ALISE Annual Conference, San Antonio, Texas, January 16–19, 2007.

Submitted August 14, 2006; accepted for publication October 17, 2006, pending revision; revision submitted December 12, 2006, and accepted for publication.

Approaches to Cross-lingual Information Access

The development of cross-lingual thesauri, subject heading lists, and name authorities, as well as the translation of the Dublin Core (DC) metadata scheme into many different languages, is ongoing. In addition to the activities of the DC Metadata Initiative for developing multilingual DC metadata, various approaches to building cross-lingual knowledge organization schemes have been developed with an eye to better access to multicultural and multilingual resources.[5]

Language engineering and linguistics communities have developed lexical tools for cross-lingual resource discovery; these include machine translation, ontology, information extraction, text summarization, and speech processing. Multilingual information resource discovery tools such as concept-based ontology (e.g., EuroWordNet and Global WordNet Association) also have been developed.[6] OLAC has been engaged in archiving, disseminating, and preserving language-culture related resources by developing the OLAC Metadata standard, which defines the format used for the interchange of metadata within the framework of the Open Archives Initiative (OAI).[7] The metadata set is based on the complete set of DC metadata terms, but the format allows for the use of extensions to express community-specific qualifiers.

In library communities, cataloging and metadata standards have been internationalized. Cross-lingual subject access via conceptual mapping of Library of Congress Subject Headings (LCSH) and cross-lingual name access through cross-linking of Library of Congress (LC) name authorities have been undertaken. The following sections present a literature review and identify the challenges inherent in transliteration and word segmentation in nonroman scripts, with particular attention to Korean. Challenges in building subject heading and name authority files for cross-lingual information access also are discussed.

Cross-lingual Subject Access: Conceptual Mapping Mechanisms

Heiner-Freiling reported the results of a survey of national libraries on subject headings conducted under the auspices of the International Federation of Library Associations and Institutions (IFLA).[8] According to the survey data, LCSH is predominantly used in twenty-four national libraries of English-speaking countries; in addition, a translated or modified version of LCSH is being used in twelve other countries. Several authors have written on the problems caused by translated subject headings across languages and cultures.[9]

Subject headings of Korean collections in North American libraries are based largely on LCSH, a translation from the source language (i.e., Korean) into LCSH in English. This author presented an earlier analysis of the problems in subject headings translated between English and Korean.[10] Problems that occur in translated subject headings likewise can be expected to occur in any metadata mapping process between the two languages.[11]

The concepts of LCSH are formulated into various syntactic forms—single noun, compound noun, noun phrase, and inverted phrase. The concept of a heading can be expressed in several different forms, leading to potential complexities and inconsistencies. Partially due to the multiple morpho-syntactic forms used in expressing the same concept, cataloger inconsistencies exist even when working with a single language, such as the assignment of subject headings in English by an English-speaking cataloger to works in the English language. The translation process between two languages only exacerbates such inconsistencies.

Korean subject cataloging suffers from the inevitable drawbacks of assigning Korean concepts by employing English subject headings. The conceptual mismatch and difficulties of translation from one language to another are largely due to different linguistic structures and socio-cultural norms. In the case of English and Korean, these structural differences are considerable, unlike between English and Spanish, because English and Korean are unrelated languages. For example, Korean is an agglutinative language in which functional particles, such as case markers and functional affixes, are attached onto the content words as grammatical operators. On the other hand, English and Spanish lack such characteristics. Instead, they are heavily dependent on word order to designate grammatical function. The manner of conveying a semantic concept may be manifested differently in Korean and English language users. Such differences in conceptual manifestation are greatly increased in the process of translation.

The following example of a translated subject heading exemplifies these problems. The romanized Korean compound phrase *Hanguk mal* could be translated as:

A Korea language/The Korean language
The language of Korea/language of Korea
Korea and a language/Korea and languages

The Korean heading may be translated into English with various forms. Major differences among these possible headings include the following: the prepositional phrase *The language of Korea* and the conjunctional phrase *Korea and languages* show indefinite and definite article variants (a versus the) and inflectional variants (language versus languages). Written Korean employs grammatical devices

such as particles (e.g., case markers denoting subject and object) and suffixes. These can be omitted in the spoken form without causing any communicational ambiguities. In the written language, as with *Hanguk mal* in the previous example, the omission of such functional words readily gives rise to ambiguity: *Hanguk ui mal* is translated as the prepositional phrase *The language of Korea*. On the other hand, *Hanguk kwa mal* is translated as *Korea and languages*. Thus, omission of the grammatical particles ui "of" and kwa "and" creates conceptual ambiguity.

Kwasnik and Rubin examined challenges in conceptual translation of classification schemes across languages and cultures.[12] They assessed differences in kinship terms in fourteen languages, revelating the challenges and problems inherent in the process of translation of a classification system. As a framework for culturally sensitive classification translation, certain modifications to the classification system (adding or deleting terms or both) reflect individual linguistic and cultural characteristics and are inevitable. In the case of one-to-two mapping, creation of cross-references is a practical step forward in clarification. In a similar manner, the use of modifiers or scope notes in order to avoid conceptual ambiguity would be advisable.

Multilingual Access to Subjects: Cross-linking Mechanisms

To date, the major project on multilingual subject headings has been Multilingual Access to Subjects (MACS), which aims at providing English, French, and German subject access in library catalogs through cross-linking techniques. Clavel-Merrin, MacEwan, and Landry reported on this project.[13] The project has been conducted by European national libraries—the Swiss National Library, the Bibliothèque nationale de France, The British Library, and Die Deutsche Bibliothek—through international collaboration under the auspices of the Conference of European National Librarians.[14]

The cross-linking technique is based on conceptual mapping among the authorized headings of three subject lists: English—LCSH, French—RAMEAU (Répertoire d'autorité matiere encyclopédique et alphabétique unifié) and German—SWD/RSWK (Schlagwortnormdatei/Regeln für den Schlagwortkatalog). Through a manual cross-linking process, conceptually equivalent linking is established. If no equivalent concept exists across the three subject headings, the heading stands alone.

The project began with a subset of headings in the areas of theater and sports. The rationale for selecting those areas was to test universality and cultural variation. The area of sports would be expected to have a high conceptual correspondence across the three languages and the three subject heading lists because the area of sports is considered to be a less culture-bound domain; conversely, the area of theater reflects culture-specific terms and concepts and low correspondence across these subject headings would be expected.

As expected, cross-linking in the area of sports yielded a high degree of equivalence. MacEwan reported that when comparing terms in a sample of 278 sports subject headings, 86 percent of headings matched across all three subject headings lists, 8 percent of headings matched across two lists, and 6 percent of headings were unmatched.[15] In the more culture-bound domain of theater, the cross-linking match was much lower than in the less culture-bound domain of sports. MacEwan reported that, when comparing terms in a sample of 261 theater subject headings, 60 percent of headings matched across all three subject heading lists, 18 percent matched across two lists, and 22 percent of headings were unmatched.[16]

A concept realized as a word in one language can be equivalent to a linguistic morpheme (the smallest unit of meaning in oral and written language), word, phrase, or clause in other languages. Thus, syntactic variations are expected to hinder the mapping process. MacEwan gave an example of the challenge seen in creating a conceptual linking system across three subject headings (English, French, and German) in the following: "*Track athletics—Coaches* in LCSH matches with *Leichtathletiktrainer* in the SWD, but in RAMEAU it is only matched by adding a subdivision to the authority record at the point of indexing a document: *Athletisme-Entraineurs.*"[17] To alleviate mapping problems caused by such syntactic variations, links between headings and strings are allowed. In addition, the creation of new headings is allowed to create a conceptual mapping between the subject heading lists, as long as there is literary warrant in the catalog of the user institution.

Conceptual Mismatch between Target and Source Languages

The conceptual mapping process is analogous to translating two or more different languages. Figure 1 illustrates some possible conceptual mismatches in the process of semantic mapping between two languages. Precise and equivalent mapping between two languages in translation does not exist. The first and second diagrams in figure 1 illustrate the necessity for strategies to deal with inexact equivalence in the case of one-to-many and many-to-one mapping. In the case of no conceptual equivalence, shown in the third diagram, the general concept in the target language might serve as an alternative for semantic mapping. However, due to the lack of specificity, the alternative general concept may not contain the original source concept, resulting in an unavoidable limitation in cross-linguistic situations.

Owing to the dramatically different language structures and cultural bases of Korean and English, translated subject headings involving these languages frequently are not equivalent to the concept of the original heading. The concept of the translated headings is either overly broad or the headings do not retain the original meaning. Thus, a more thorough analysis and understanding of the very different Korean and English language structures are needed to alleviate this inevitable difficulty.

The subject heading that follows, taken from a MARC record describing a Korean monograph, *pumasi,* illustrates the challenges faced in conveying the original concept in the process of mapping from the Korean concept of a word to LCSH.

650 0 Interpersonal relations.
651 0 Kyonggi-do (Korea)$x social life and customs.

The title of the book is *pumasi* (exchange of services/labor) *wa* (and) *chong* (affection) *ui* (of) *ingan* (human) *kwangye* (relationship). The translation could be *The interpersonal relations of the exchange of labor and affection*. The word pumasi describes the social structure of Korea in the agricultural context. The pumasi is the system by which people effectively provide help to one another. People who are in need can obtain financial and other help from others for a short period without paying interest. They will return the pumasi on some other occasion when the people who gave help are themselves in need of help. This system was originally developed in a traditional agricultural society and then transferred into the urban society of modern Korea. The underlying concept of pumasi may be stated thus: *solidarity with affection in a community*.

LCSH does not have a heading that is equivalent to the pumasi system. This is because pumasi is a product of Korean culture. In order to denote the subject heading, then, a broad and general heading such as *social life and customs* would be employed for this monograph in the topical subdivision of the heading (i.e., 651). As can be seen, the translated subject heading in the above record loses the original concept of the Korean heading due to conceptual mismatch.

Cross-lingual Name Access through Cross-linking Mechanisms

Two major projects on cross-lingual name access through the cross-linking mechanism utilizing roman script currently are employed. One is the Virtual International Authority File (VIAF), a joint project between LC and Die Deutsche Bibliothek, with OCLC's research support.[18] VIAF is a single personal name authority file that combines the name authority files of both institutions through the cross-linking mechanism.

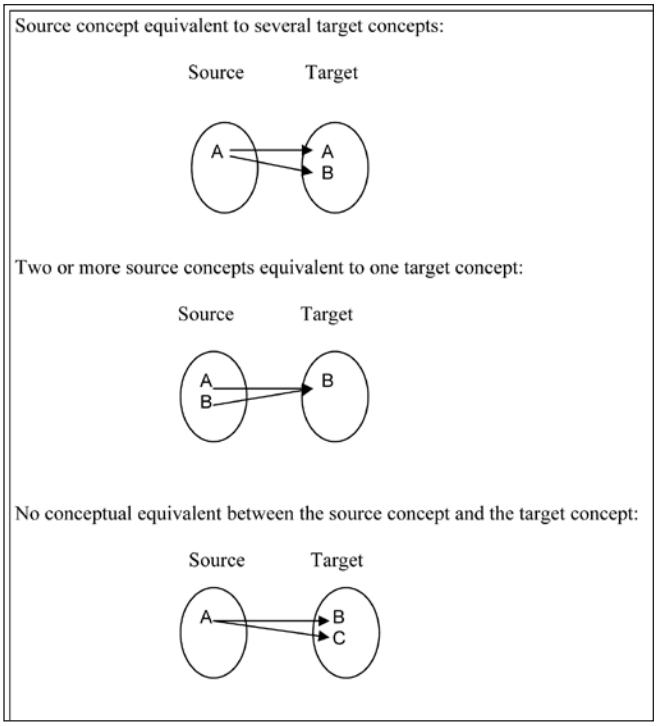

Figure 1. Conceptual equivalence

Source: Jung-ran Park, "Hindrances in Semantic Mapping among Metadata Schemes: A Linguistic Perspective," *Journal of Internet Cataloging* 5, no. 3 (2002): 74.

In the VIAF project, the authority records from Die Deutsche Bibliothek are matched to the corresponding LC authority records through the cross-linking mechanism. Following this linking process, maintaining the authority files and providing user access to the files will be through the shared OAI servers. Upon the completion of the project, each user group in the United States or Germany will be able to view personal name records established by the other institution and view the personal name records of each user group's own language.

The other project dealing with roman script is Linking and Exploring Authority Files (LEAF), which was established in 2001 with the involvement of fifteen organizations utilizing eight languages.[19] Clavel reported two principal challenges in establishing a cross-lingual authority file.[20] Both challenges are derived from linguistic variation and ambiguities across languages. First are language-specific features such as the order of components in compound names, location of particles, and numbering system for kings and popes. The second challenge concerns standardization of methods for disambiguation of homonyms. Natural lan-

guage is full of lexical ambiguities. For instance, homonymy creates ambiguity (e.g., bank [building] versus bank [river]). Homonyms have the same lexical form but manifest unrelated meanings that are arbitrarily developed. The *Anglo-American Cataloguing Rules,* 2nd edition (*AACR2*) chapters 22 through 26 present pragmatically constrained disambiguation techniques for the names of persons, corporate bodies, and places by differentiating contexts.[21] For example, to disambiguate identical names, birth or death dates (or both) are added (e.g., John Q. Smith [1904–1972] versus John Q. Smith [1905–]). In the case of ambiguous corporate body names, a qualifier is added—e.g., John Smith (firm). According to Clavel, the addition of academic and nobility titles is generally standardized for disambiguating homonyms.[22] The specification of profession or activity, however, is much less standardized. Accordingly, this creates problems in cross-linking of authority files across languages.

Several authors have looked at nonroman scripts (particularly East Asian languages such as Korean, Japanese, and Chinese) and have found that transliteration causes cross-lingual name access problems, because of the nature of the language.[23] Names in the Korean, Chinese, and Japanese languages utilize Chinese ideographs owing to a common history; thus, variant forms of names are represented in these languages. For example, in the case of the Korean name, *Hangul (*Korean vernacular script*)*, Chinese ideograph and the transliterated form are all used.

When discussing Korean, one must take into account the differences in transliteration schemes between those based on phonetic structure and those based on morphemic structure. Differences in transliteration schemes are also applicable to other nonroman scripts.

For instance, LC's relatively recent adoption of the Pinyin transliteration scheme from the Wade-Giles scheme in transcribing Chinese language materials illustrates the complex issues surrounding the differences in transliteration schemes even involving the same language. Arsenault reported on an experiment in retrieval efficiency among monosyllabic Pinyin, polysyllabic Pinyin, and Wade-Giles while searching known item exact title and keywords in title.[24] The findings of the study demonstrate that the polysyllabic Pinyin system, which transcribes Chinese according to syntactic unit (i.e., word by word), significantly increases retrieval efficiency compared to monosyllabic Pinyin and Wade-Giles, which share the feature of transcribing Chinese morpheme by morpheme.

Naito presented a variety of ways of transcribing the same Japanese name, such as phonetic transcription in Hiragana and phonetic transcription in Katakana, transcription in simple form, and Chinese scripts.[25] Table 1 (from Naito) illustrates this.

This author presented issues relating to the Korean transliteration scheme.[26] In South Korea, no unified transliteration scheme is used. Different transliteration schemes are employed in different sectors for varying uses. For example, libraries and publishing industries employ the McCune-Reischauer (MR) system in publication and bibliographic records.[27] The Yale system is uniformly used by linguists within Korea and abroad.[28] Lastly, government documents, including street signs and road maps, employ the Ministry of Education system.[29]

The differences among these schemes reflect the linguistic representation of sound systems. The MR system and Ministry of Education system are based on the phonetic structure of Korean. Transliteration based on phonetic structure encodes words in the manner in which they are pronounced. For example, in English the word *two* is transcribed phonetically as [*tu*].

The Yale system is based on morphemic structure. Morphemic structure-based transliteration transcribes the base form of a word regardless of sound changes. Korean is a language that employs rich morpho-phonemic complexity. 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 morpho-phonemically. For example, the form of the Korean word *mul* (water) is changed into *muri* when the subject case particle *-i* is attached to it. Morphemic structure-based transliteration is not reflective of sound change as is the phonetic type of transliteration utilized in the MR scheme; instead, it reflects the base form.

The current cataloging system dealing with Korean materials employs the MR transliteration scheme. One of the major drawbacks of the use of the MR system is that it causes semantic loss. This is especially critical in the area of name access. Transliteration of words following the way in which they are pronounced has the potential of representing

Table 1. Japanese personal name

黒澤明	Form he used
黒沢明	Simplified character of "澤"
くろさわ　あきら	Phonetic description in Hiragana
クロサワ　アキラ	Phonetic description in Katakana
ｸﾛｻﾜ ｱｷﾗ	Phonetic description in Katakana by computer half-width character still in use
Kurosawa Akira KUROSAWA Akira	Romanized form Family name + Given name
Akira Kurosawa Akira KUROSAWA	English form (?) Given name + Family name

Source: Eisuke Naito, "Names of The Far East: Japanese, Chinese and Korean Authority Control," *Cataloging & Classification Quarterly* 38, no. 3 (2004): 257.

a name ambiguously. For example, the Korean name *Kim Sok-min* becomes *Kim Song-min* according to the MR system. With author names transliterated according to the MR system, ambiguity becomes almost inevitable. The linguist Ramsey noted that "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."[30]

The MR system, based as it is on phonetic structure, does not disambiguate different meanings of homographs (i.e., same words but different meanings), one of the primary causes of semantic ambiguity. This phenomenon can be illustrated by an example in English: *two, to, too*. If these three lexical items are transcribed according to the pronunciation [*tu*], the resulting semantic ambiguity can be clearly seen. This happens frequently with the MR scheme. Such ambiguities inevitably cause significant impediments in the process of information retrieval.

In addition, the MR system results in variations in the creation of bibliographic records. When catalogers transcribe words according to pronunciation, they can create inconsistent and arbitrary records. This is based on the fact that the pronunciation of words can vary according to speech style. If a cataloger pronounces a word or phrase using careful speech style, the resulting transcription would be different from that of a transcription based on casual speech style. The creation of differing bibliographic records is thus entirely possible, either by the same cataloger or different catalogers transcribing identical material.

The following bibliographic record illustrates this problem.

100 1	Kim, Young-un,$1927-
245 10	Ceh-2 **k^onggungnon** : $bkungmin kukka ^ui wans^ong ^ul wihay^o /$cKim Yong-un.
246 3	Ch"io"an.
260	S^oul T"^ukpy^olsi :$Chisik San^opsa, $c1998.

The portion of the title field (245) in bold, *k^onggungnon,* reflects the casual speech style. If the cataloger who created this record had pronounced it using careful speech, the final consonant of the first syllable (i.e., *ko**n***) remains as a nasal sound, as indicated in bold: *k^o**n**gungnon*, as opposed to *k^o**ng**gungnon*. In casual speech, however, the nasal sound [*n*] becomes assimilated into the following velar sound [*ng*].

The MR transliteration scheme contains inherent inconsistencies that can have a significant impact on information organization and retrieval. Semantic ambiguity, inconsistency, and semantic loss are critical issues hindering information retrieval and sharing bibliographic records. Consequently, the goals of bibliographic control are not achieved.

Problems of Word Segmentation

Difficulty in word segmentation occurs in agglutinative languages such as Japanese and Korean because of their inherent morpho-syntactic flexibility. Agglutinative languages allow functional particles such as case markers and inflectional affixes to be attached onto the content words as grammatical operators. For example, the word *muli* [water + subjective case affix] is composed of the content word (i.e., *mul*: water) and the functional affix (i.e., *i*: subjective case marker). This creates flexible word segmentation between functional and content words. Such flexibility of word segmentation in Korean creates inconsistent and arbitrary practice in word division; such inconsistency can be found in even the most authoritative Korean dictionaries. According to Yi Sung-u, word segmentation errors appear in 29 percent of Korean standard books in the school system.[31] This highlights the difficulty in conducting word segmentation in the written Korean form.

Arbitrary word segmentation does not cause communication problems in everyday language use, since communicative ambiguities stemming from inconsistent word segmentation can be resolved through contextual cues. However, such flexibility in word segmentation is a critical factor in hindering information sharing and discovery in the digital environment, which does not provide contextual cues.

The Library of Congress *ALA-LC Romanization Tables* provides rules specifying word segmentation and offer four basic underlying principles.[32] The first basic principle is "Each word or lexical unit (including particles) is to be separated from other words." [33] The following Korean bibliographic record illustrates this principle.

245 00	**Y^oksa sok ^ui in'gan kwa chis^ong ^ul t"amgu handa** /$c Kim Chae-yong . . . [et al.] p"y^on.
250	Che l-p"an.
260	S^oul :$bHan'gilsa,$c1998.

The title field (245) can be segmented in the following way: *Yoksa^ sok^ ui^ in'gan^ kwa^ chisong^ ul^ t"amgu^ handa*. The segmentation is denoted by the mark ^, designating a total of eight word divisions. This principle follows one of the suggestions presented at the 1981 workshop conference on Korean transliteration, held at the University of Hawaii under the auspices of the Korean Studies Center, and reported by Austerlitz.[34] The main aim of the

conference was to examine the Korean transliteration system (i.e., MR system) to produce consistent guidelines for transliterating Korean language.

This principle creates problems when users search a bibliographic record because word division following the LC principle is not utilized by Korean users; it is contrary to conventional practices of the language. The previous example title consists of only three word divisions in the Korean written form: *Yoksasokui^ in'gankwa^ chisongul^ t"amguhanda.* Moreover, this rule presents another intrinsic difficulty. It applies only to case particles of a noun phrase, not to affixes of verb phrases. Thus, the word division principle is not applied to entire units of the sentence.

The MR transliteration scheme based on phonetic structure has critical drawbacks because it causes semantic loss, semantic ambiguity, and cataloging inconsistency. A transliteration scheme based on morphemic principles has substantial merit because it significantly contributes to resolving semantic ambiguity and inconsistency. One of the principal advantages of basing transliteration on morphemic principles is that the need for diacritical symbols also is substantially reduced, in contrast to a transliteration scheme based on phonetic principles, which increases the employment of diacritical symbols.

Word segmentation in agglutinative languages is very flexible. Even though guidelines and rules for word division exist, inconsistent and arbitrary practices are inevitable. An automatic parser of word segmentation based on linguistic principles is critically needed to ensure consistency of bibliographic records.

Linguistic Universality and Relativity across Language Structures

Impediments to enhancing access to cross-cultural and cross-lingual resources are largely derived from the complexities and variation of linguistic structures across languages. Linguistic and cultural approaches in developing cross-lingual and cross-cultural knowledge organization systems are critically needed.

The facility of natural language, in all its complexity, variability, and richness, is the defining aspect of humanity. This very complexity of expression and richness of lexicalization and linguistic structures becomes problematic in the electronic environment of information retrieval. Even though natural language possesses some characteristics that are independent of a specific language, many more language-specific characteristics exist. Such language-specific characteristics demonstrate that the structure of language is so closely intertwined with its source culture and society that it is inseparable from it. Natural language is not just mere arrangements of words, but the mirror of culture. Combinations and arrangements of words do not reflect specific cultural and pragmatic meanings that are inherent characteristics in any given language structure.

Language-specific variations and differences in lexicalization patterns can be found easily in everyday language uses such as naming conventions, kinship terms, address forms, numbering systems, color terms, and names for body parts. For example, in Anglo-American society, building designations (e.g., *LeBow* College of Business), brand names (e.g., *Ford*), and even common reference nouns (e.g., *maverick, boycott, lynch*) originating from family names or titles are common. Conversely, this phenomenon is nonexistent in Korean language and society. Thus, one can say that this English-specific naming convention manifests the cultural trait of Anglo-American society.

Collectivist-oriented cultural and social norms, based on hierarchical structure, are closely reflected in the Korean language. This can be especially seen in the sophisticated honorific system and in the employment of various linguistic devices, such as lexical items existing in both plain and honorific form (e.g., *na/cho* [plain/honorific form] 'I', *nai/yonsey* [plain/honorific form] 'age', *chada/chumusida* [plain/honorific form] 'sleep: verb), to name a few. It is also seen in syntactic structures (e.g., honorific agreement in subject/object, predicate, and case markers). Such variant lexical forms are merely one illustration of a synonymy phenomenon that is not found in English, as shown in table 2.

The Need to Develop Interoperable Guidelines for Cross-linking Names and Subjects and Conceptual Mapping

A critical need for the development of common guidelines for cross-linking of names (e.g., person, place, corporate body) across languages exists. Development of such interoperable cross-linking guidelines should be guided by the examination of morpho-syntactic variations across language structures, especially for the structures of names.

Word segmentation and transliteration schemes dealing with nonroman scripts also play a part in limiting access to cross-lingual and cross-cultural resources. Standardization of such transliteration schemes and development of mechanisms geared toward consistent word segmentation also are critically needed. Specifically, reexamination of transliteration schemes and development and application of a morpho-syntactic parser based on linguistic principles for automatic word segmentation are vital conditions for cross-lingual information access.

Development of knowledge organization schemes for cross-lingual subject access also is hindered by the lack of common conceptual mapping criteria that are interoperable across languages and cultures. Semantic mapping, involv-

Table 2. Korean lexical honorific system

English equivalent	Plain form	Honorific form
age: noun	*nai*	*yonse*
house: noun	*chip*	*taek*
sleep: verb	*chada*	*chumusida*
eat: verb	*mokta*	*tusida*

Source: Jung-ran Park, "Hindrances in Semantic Mapping among Metadata Schemes: A Linguistic Perspective," *Journal of Internet Cataloging* 5, no. 3 (2002): 63.

ing metadata and subject heading lists across languages, is one of the most critical issues in resource discovery and information exchange. Without achieving interoperability of semantic mapping, application of cross-lingual knowledge organization tools for the retrieval of networked resources will be significantly hindered. In order to develop interoperable conceptual mapping guidelines across languages and cultures, identification of lexicalization patterns based on semantic, syntactic, and pragmatic linguistic analysis is critically needed.

Cross-linguistic differences result in conceptual and lexical gaps and overlaps between target and source languages that present themselves during the mapping process. Conceptual mapping between languages presents a variety of lexical gaps and overlaps including inexact equivalence, partial equivalence, nonequivalence, and single-to-multiple equivalence. Culture-specific language characteristics suggest that, in order to overcome problems in the development of cross-lingual knowledge organization tools (e.g., subject headings, thesauri, metadata) and to ensure interoperability among these tools cross-linguistically, language-specific characteristics must be taken into account.

Conclusion

Complexities and variations of linguistic structures across languages and cultures have a significant effect on name and subject access across languages. Thus, study of linguistic and cultural approaches to developing cross-lingual and cross-cultural knowledge organization systems is critically needed. The major research gaps in current literature concern addressing issues in relation to developing interoperable guidelines for cross-linking of names and developing common conceptual mapping criteria that are interoperable across languages and cultures for cross-lingual subject access.

This underlies the necessity of future studies in morpho-syntactic variation across languages for cross-lingual name access and an examination of lexicalization patterns based on semantic, syntactic, and pragmatic linguistic analysis for cross-lingual subject access. Drawbacks in word segmentation and transliteration schemes dealing with nonroman languages also call for reexamination of transliteration schemes and for the development of a morpho-syntactic parser for automatic word segmentation.

References

1. International Society for Knowledge Organization, "Knowledge Organization and the Global Information Society," 8th International IKSO Conference. www.ucl.ac.uk/isko2004/index.htm (accessed Aug. 4, 2006); J. Williamson and Clare Beghtol, eds., "Knowledge Organization and Classification in International Information Retrieval," *Cataloging & Classification Quarterly*, 37 no. 1/2 (2003); Alvaro Quijano-Solis, Pilar Maria Moreno-Jimenex, and Reynaldo Figueroa-Servin, "Automated Authority Files of Spanish-Language Subject Headings," *Cataloging & Classification Quarterly* 29, no. 1/2 (2000): 209–23; Reynaldo D. Figueroa-Servin and Berta Enciso, "Subject Authority Control at El Colegio de Mexico's Library: The Whats and Hows of a Project," *Cataloging & Classification Quarterly* 32, no. 1 (2001): 65–80; Jung-ran Park, "Hindrances in Semantic Mapping among Metadata Schemes: A Linguistic Perspective," *Journal of Internet Cataloging* 5, no. 3 (2002): 59–79; Barbara H. Kwasnik and Victoria L. Rubin, "Stretching Conceptual Structures in Classifications across Languages and Cultures," *Cataloging & Classification Quarterly* 37, no. 1/2 (2003): 33–47; Zahiruddin Khurshid, "Arabic Script Materials: Cataloging Issues and Problems," *Cataloging & Classification Quarterly* 34, no. 4 (2002): 67–77; Annarita Sanso, "Ancient Italian State: An Authority List Project," *Cataloging & Classification Quarterly* 39, no. 1/2 (2004): 577–84.
2. John Agada and Brenda Hough, "The Emporia-Nigeria Project: Building Global Partnerships to Support Development of Civil Society in Nigeria," unpublished paper presented at the 2005 ALISE Annual Conference, Jan. 11–14, 2005, Boston; Sergio Chaparro-Univazo, "Some Issues on LIS Education and Collaboration in Latin America," paper presented at the 2005 ALISE Annual Conference, Jan. 11–14, 2005, Boston. http://dlist.sir.arizona.edu/701 (accessed Mar. 24, 2006); Bharat Mehra, "Cross-Cultural Perspective of International Doctoral Students: 'Two-way' Learning to Further Internationalization in LIS Education," unpublished paper presented at the 2005 ALISE Annual Conference, Jan. 11–14, 2005, Boston.
3. Cognitive Science Laboratory, Wordnet: A Lexical Database for the English Language. http://wordnet.princeton.edu (accessed Dec. 1, 2006); EuroWordNet. www.illc.uva.nl/EuroWordNet (accessed Dec. 1, 2006).
4. Christiane Fellbaum, ed., *WordNet: An Electronic Lexical Database* (Cambridge: MIT Pr., 1998); George A. Miller et al., "Introduction to WordNet: An On-Line Lexical Database," *International Journal of Lexicography* 3, no. 4 (1990): 235–44; Eduard Hovy et al. *Multilingual Information Management: Current Levels and Future Abilities.* A report commissioned by the U.S. National Science Foundation and also delivered to the European Commission's Language Engineering Office and the U.S. Defense Advanced Research Projects Agency, Apr. 1999. www-2.cs.cmu.edu/~ref/mlim/index.shtml (accessed Aug. 4, 2006); Piek Vossen, "EuroWordNet: Building a Multilingual Database with Wordnets for European Languages," *ELRA Newsletter* 3, no. 1 (1998): 7–10.
5. Jung-ran Park, "Language-Related Open Archives: Impact on Scholarly Communities and Academic Librarianship," *E-JASL: The Electronic Journal of Academic and Special Librarianship* 5, no. 2/3 (2004), http://southernlibrarianship.icaap.org/content/v05n02/park_j01.htm (accessed Aug. 4, 2006); Jung-ran Park, "Human Language: Resources from Linguistics and Beyond," *College & Research Libraries News* 66, no. 3 (2005): 173–76, 228.
6. EuroWordNet, www.illc.uva.nl/EuroWordNet (accessed Mar. 24, 2006); Global WordNet Association, www.globalwordnet.org (accessed Mar. 24, 2006).
7. OLAC: Open Language Archives Community. www.language-archives.org (accessed Aug. 4, 2006); OLAC Metadata [standard], www.language-archives.org/OLAC/metadata-20070405.html (accessed Mar. 24, 2006).
8. Magda Heiner-Freiling, "Survey on Subject Heading Languages Used in National Libraries and Bibliographies," *Cataloging & Classification Quarterly* 29, no. 1/2 (2000): 189–98.
9. Quijano-Solis, Moreno-Jimenex, and Figueroa-Servin, "Automated Authority Files of Spanish-Language Subject Headings," 209–23; Figueroa-Servin and Enciso, "Subject Authority Control at El Colegio de Mexico's Library," 65–80; Park, "Hindrances in Semantic Mapping," 59–79; Kwasnik and Rubin, "Stretching Conceptual Structures," 33–47; Khurshid, "Arabic Script Materials," 67–77; Sanso, "Ancient Italian State," 577–84.
10. Park, "Hindrances in Semantic Mapping," 59–79.
11. Jung-ran Park, "Semantic Interoperability across Digital Image Collections: A Pilot Study on Metadata Mapping," in CAIS/ACSI 2005 *Data, Information, and Knowledge in a Networked World,* ed. Liwen Vaughan. Proceedings of the 2005 Annual Conference of the Canadian Association for Information Science held with the Congress of the Social Sciences and Humanities of Canada (University of Western Ontario, London, Ontario, June 2–4, 2005). www.cais-acsi.ca/proceedings/2005/park_J_2005.pdf (accessed Mar. 24, 2006).
12. Kwasnik and Rubin, "Stretching Conceptual Structures," 33–47.
13. Genevieve Clavel-Merrin, "MACS (Multilingual Access to Subjects): A Virtual Authority File Across Languages," *Cataloging & Classification Quarterly* 39, no. 1/2 (2004): 323–30; Andrew MacEwan, "Crossing Language Barriers in Europe: Linking LCSH to Other Subject Heading Languages," *Cataloging & Classification Quarterly* 29, no. 1/2 (2000): 199–207; Patrice Landry, "Multilingual Subject Access: The Linking Approach of MACS," *Cataloging & Classification Quarterly* 37, no. 3/4 (2004): 177–91.

14. MACS: Multilingual Access to Subjects. https://macs.vub.ac.be/pub (accessed Mar. 24, 2006).
15. MacEwan, "Crossing Language Barriers in Europe," 199–207.
16. Ibid.
17. Ibid., 203.
18. OCLC Research Projects, "Virtual International Authority File." www.oclc.org/research/projects/viaf (accessed Aug. 4, 2006).
19. Institute of Information Systems & Information Management, Joanneum Research, "LEAF—Linking and Exploring Authority Files," www.joanneum.at/index.php?id=358&L=1&L=1 (accessed Mar. 24, 2006); Jutta Weber, "LEAF: Linking and Exploring Authority Files," *Cataloging & Classification Quarterly* 38, no. 3/4 (2004): 227–36.
20. Pierre Clavel, "LEAF," paper presented at ELAG 2003: Cross Language Applications and the Web (27th Library Systems Seminar, Switzerland, Apr. 2–4, 2003). http://elag.kb.nl/elag2003/www.elag2003.ch/pres/pres_clavelp.pdf (accessed Mar. 24, 2006).
21. *Anglo-American Cataloging Rules,* 2nd ed., 1998 rev. (Ottawa: Canadian Library Assn.; London: Library Assn. Publishing; Chicago: ALA, 1998).
22. Clavel, "LEAF."
23. Eisuke Naito, "Names of the Far East: Japanese, Chinese, and Korean Authority Control," *Cataloging & Classification Quarterly* 38, no. 3 (2004): 251–68; Jung-ran Park, "Information Retrieval of Korean Materials Using the CJK Bibliographic System: Issues and Problems," in *Proceedings of the Second KSAA Biennial Conference: Korean Studies at the Dawn of the Millennium,* Young-A Cho, ed. (Monash University, Melbourne, Australia, Korean Studies Association of Australasia, Sept. 24–25, 2001), 245–55. www.arts.monash.edu.au/korean/ksaa/conference/21jungranpark.pdf (accessed Mar. 24, 2006); Hee-sook Shin, "Quality of Korean Cataloging Records in Shared Databases," *Cataloging & Classification Quarterly* 36, no. 1 (2003): 55–90; Jian Wang, "Chinese Serials: History, Characteristics, and Cataloging Considerations," *Cataloging & Classification Quarterly* 36, no. 1 (2003): 41–54.
24. Clément Arsenault, "Word Division in the Transcription of Chinese Script in the Title Fields of Bibliographic Records," *Cataloging & Classification Quarterly* 32, no. 3 (2001): 109–27.
25. Naito, "Names of the Far East," 251–68; Park, "Information Retrieval of Korean Materials."
26. Park, "Information Retrieval of Korean Materials," 245–55.
27. George M. McCune and Edwin O. Reischauer, "The Romanization of the Korean Language: Based upon Its Phonetic Structure," *Transactions of the Korea Branch of the Royal Asiatic Society* 29 (1940): 1–55.
28. Yale Romanization: Encyclopedia. http://experts.about.com/e/y/ya/Yale_Romanization.htm (accessed Aug. 4, 2006); Yale System of Korean Romanization. www.tufs.ac.jp/ts/personal/choes/Eyale.html (accessed Aug. 4, 2006).
29. Ministry of Culture and Tourism of the Republic of China, *The Romanization of Korean* (Seoul: Ministry of Culture and Tourism, 1979); Munhwa Kangwangbu kosi (The Ministry of Culture and Tourism of the Republic of Korea Announcement). Ministry of Culture and Tourism of the Republic of China, *Sae Romacha pyogipop* (New System of Romanization for the Korean Language, 2000).
30. S. Robert Ramsey, "Writing Korean with Roman Letters," *Korean Journal* 22, no. 8 (1982): 33.
31. Sung-u Yi, *Sae Machumpop kwa Kyojong ui Sirches* (Seoul: Omungak, 1993).
32. Library of Congress and the American Library Association, *ALA-LC Romanization Tables: Transliteration Schemes for Non-Roman Scripts*, (Washington, D.C.: Cataloging Distribution Service, 1991).
33. Ibid., 82.
34. Robert Austerlitz et al., "Report of the Workshop Conference on Korean Romanization," *Korean Studies* 4 (1980): 111–25.