Tadao Miyamoto, Joseph F. Kess & Tomoyoshi Inoue

Cross-Linguistic Mental Representations of Katakana Loanwords in Japanese

Series B: Applied and Interdisciplinary Papers
ISSN 1435-6481
Essen: LAUD 2000 (2., unveränderte Auflage 2006)
Paper No. 287

Universität Duisburg-Essen
Tadao Miyamoto, Joseph F. Kess and Tomoyoshi Inoue

University of Victoria (British Columbia) & Doshisha University (Kyoto)

Cross-Linguistic Mental Representations
of Katakana Loanwords in Japanese

Copyright by the author
2000 (2., unveränderte Auflage 2006)
Series B
Applied and Interdisciplinary
Paper No. 287

Reproduced by LAUD
Linguistic Agency
University of Duisburg-Essen
FB Geisteswissenschaften
Universitätsstr. 12
D- 45117 Essen

Order LAUD-papers online: http://www.linse.uni-due.de/linse/laud/index.html
Or contact: laud@uni-due.de
Cross-linguistic Mental Representations of Katakana Loanwords in Japanese

1. Introduction

This paper is part of an ongoing research project on Japanese psycholinguistics. The main aim of our project has been to produce a comprehensive overview of Japanese attempts to probe how natural language is produced, comprehended, stored, and recalled (see Kess & Miyamoto, 1994). When we surveyed the rich field of Japanese psycholinguistics, we quickly realized that the most fertile areas for research has been the area of kanji and kana processing, and that one of the longest and richest traditions in modern Japanese psycholinguistics has been in the area of lexical access and word recognition, and the way in which the four orthographic types (katakana, hiragana, romaji, and kanji characters) affect processing. Hence, in the first half of this paper, we will provide an overview of the theoretical issues in word recognition and the way in which the Japanese mental lexicon reflects processing strategies which are tied to the quadripartite Japanese orthographic system (see Kess & Miyamoto, 1999). In the second half of the paper, we will focus our attention on katakana processing, discussing the results of our pilot study on the acquisition of katakana syllabary by learners of Japanese.

2. Japanese Orthographic System

Orthographic systems are universally classified as one of three types. Alphabetic systems like English, Greek, and Russian employ symbols for individual phonemes, syllabic systems like Cherokee and Miao employ symbols for syllables, and logographic systems like Chinese employ symbols for morphemes or words (see Gelb, 1963). The Japanese orthography is unique in that it employs all three types of orthography: logographic kanji, syllabic kana, and alphabetic romaji.

Kanji were imported into Japanese orthography from Chinese. Unlike Chinese or Korean, Japanese kanji characters can have two possible types of reading for a given kanji; Chinese on-readings can compete with Japanese kun-readings of the same kanji. Multiple readings for individual kanji are also common, with one kun-reading and one or more on-readings of the character because Japanese borrowed kanji in separate and distinct historical periods, complicating reading of kanji in the modern writing system. In terms of graphic configuration, kanji characters can be 'simple' or they can be 'complex'. A simple character is one that can stand alone. A certain number of these may also be used as a 'radical' which forms part of a larger, more complex single character. Among complex kanji, the largest number are phonetic-semantic kanji (keisei moji), containing a semantic 'radical' which gives a rough approximation of where the kanji fits categorically and a 'phonetic' which may suggest the Chinese reading for the kanji.
Japanese text is never limited to just kanji, but is intermixed with kana syllabary symbols in what is known as kanji/kana majiribun 'kanji/kana mixed sentences'. There are two types of kana syllabaries, katakana and hiragana, both of which are derived from kanji simplifications. Hiragana and katakana scripts represent sound units which corresponds to a mora, for which the canonical shape is a Consonant-Vowel (CV) combination. The grapheme-to-mora correspondence ratio is extremely high for kana in both syllabary systems. Hiragana and katakana share the same syllabic reference point, so that the same syllable can be transcribed by either system. However, the katakana syllabary, which is more angular in the shape of its symbols, is commonly taken as the appropriate transliteration medium for loanwords into Japanese from other languages. The trend toward internationalization and the constant interaction with Western culture in Japan has seen an enormous increase in the loanword inventory of the Japanese lexicon, and these loanwords are regularly spelled in katakana. Hiragana, which is more cursive in the shape of its symbols, is most commonly used in writing the non-content words and grammatical morphemes not usually presented by kanji characters. It is used for morphological endings, function words, and the rest of the grammatical scaffolding of Japanese sentences. Kanji may roughly constitute a third of the total number of symbols in the average sentence (see Saito, 1981), but what is not kanji is usually hiragana. The increasing use of the Latin alphabet-based romaji constitute fourth script type. Romaji is used to directly represent loanwords with no transliterative intervention of the katakana syllabary. The use of romaji is abundant in printed advertisements, especially those aiming at younger consumers.

As a result, Japanese has a mixed orthographic system, employing two syllabaries which match the relatively simple syllabic structure of the language, incorporation of romaji, and a large inventory of logographic kanji which are used mainly to represent content words. The issues in lexical access for written Japanese words are thus complicated by the fact that Japanese does not have a single script type (see Kess & Miyamoto 1999, for details).

3. Kanji Processing

Current psycholinguistics is quite preoccupied with the structure of the mental lexicon; so also is the Japanese psychological literature, although largely through the medium of kanji processing (see Kess & Miyamoto, 1999). The mental representation for lexical items presented through kanji invites a number of processing questions. One processing question involves the role of phonological factors: can meanings of words written in kanji be understood even when their phonetic codes are not retrieved from the written transcriptions? A second key question is whether the 'semantic' qualities of certain radicals are in any way employed during the semantic interpretation of kanji words? The third question has to do with the fact that Japanese words are typically compounds, and not words represented by single kanji. The issue is whether compound kanji are recognized and processed as integrated units, or whether their successful recognition is contingent upon the processing of their individual kanji components?

Some suggest that there is a 'universal phonological principle' inherent in accessing the mental lexicon: whichever writing system is used, be it alphabetic, syllabic, or logographic, lexical access ultimately invokes the phonological characteristics of the word stored away (Perfetti & Zhang, 1995). With respect to kanji processing, one influential view, the speech recoding view, claims that character processing in lexical access automatically proceeds
from the written form of the word through the speech coding for the word. The idea is that activation of the phonological properties of a kanji word is an automatic and integral component in the path of accessing the word's identity in the mental lexicon. And there have been various forms of support for this hypothesis (see Kess & Miyamoto, 1999, and references cited therein). These studies, however, have not definitively answered the question of when that phonological information is made available. Does it take place at the pre-lexical stage before word recognition has been achieved? Or at the post-lexical stage once word recognition has been achieved? Or is it an at-lexical event as an integral part of the word recognition event. (The problem of phonological activation in Japanese is complicated by the problem that there is usually more than one reading for a given kanji.) One wonders if future research will provide a definitive answer about the time course for phonological information.

A key question with processing of semantic information in kanji processing is: do readers of kanji process at some unconscious level the semantic information contained in the radicals embedded within complex kanji? (Radicals are larger than individual strokes, and are rather those stroke clusters which can be considered the identifiable structural pieces which make up the kanji.) The answer to this question is important for a universal theory of word recognition, for it is the Japanese version of the debate regarding analytic decomposition of morphologically complex words in alphabetic languages like English. The closest analogy to the issue of prefix-stripping in English words such as relish vs. retrain may be the Japanese complex kanji. Does analytic decomposition of complex kanji take place in Japanese, and if decomposition takes place, is such analysis optional or obligatory for the reader? Such questions about kanji have most often been probed with semantic priming tasks (see Ohta, 1991, for a review of the literature on direct priming). Priming experiments which examine lexical access procedures for complex kanji offer some interesting perspectives on the full-listing vs. decompositional/affix-stripping positions on how word recognition is supposed to take place (see Kess, 1992). But like the evidence for decomposition in languages like English, the evidence for this is mixed as regards Japanese kanji (see Kess & Miyamoto, 1999, for details).

The focus on individual kanji, simple or complex, does not necessarily illuminate the cognitive procedures employed in dealing with many common Japanese words because many common words in Modern Japanese are represented by a compound jukugo, a polysyllabic composite of two to four kanji. A key question with respect to processing of compounds is whether their lexical access is performed at the whole-word level or the individual character level. Like the evidence for compounds in languages like English, the evidence for lexical access procedures is mixed as regards Japanese kanji compounds. It does seem certain that the phonological rendering of a kanji is highly dependent on the intra-word context, and is then finalized at the word level. However, there is also some evidence showing that since the appropriate reading of kanji characters in a kanji compound is ultimately decided by the whole-word context, whole-word translation from orthography to phonology is the most dependable processing procedure. And whole-word access seems to outweigh sub-lexical considerations which would access probabilistic values for readings of the constituent kanji (see Kess & Miyamoto, 1999).
4. Kana Processing

Not only kanji processing, but kana processing has also received a fair amount of research interests (see Kess & Miyamoto, 1999). With respect to the nature of kana characters, some studies suggest that the property of the phonological representations may be different for scripts based on alphabets vs. syllabaries. Perhaps syllabaries function midway between words and alphabetic letters, hence representing higher level processing units than alphabetic letters (Besner, 1990; Shimomura & Yokosawa, 1995). Tamaoka and Taft (1994), however, found that the smallest unit of phonological processing in Japanese is the phonemic segment, as in the alphabetic writing system. They modified katakana words for a lexical decision task in which words like ka-me-ra 'camera' could appear in three altered shapes: ko-me-ra, so-me-ra, and so-ki-ra. Subjects were presented with 30 such stimulus sets for a lexical decision task. Response times required were longer for the ko-me-ra type stimuli, suggesting that subjects were engaged longer in seeking a lexical address for this type which, unlike the others, differed from their originals only in single phonemes. Although the mora is the smallest unit of orthographic representation, Japanese subjects were obviously sensitive to phonemic segments in processing these kana words, providing evidence that phonemic segments are relevant to phonological processing of kana representations.

With respect to lexicality of kana words, there is some debate on whether recognition of kana words is decided mostly at lexical level (see Kess & Miyamoto, 1999). There is some support for the significance of lexicality. Yamada (1992) constructed hiragana non-words as a control measure in an experiment contrasting kana readings of common words. His finding was that orthographically unfamiliar hiragana words were named faster than orthographically unfamiliar hiragana non-words. This is a hiragana mirror image of Besner and Hildebrandt's (1987) findings of the same lexicality effect for katakana words vs. katakana non-words. Taken together, these findings for hiragana and katakana confirm that lexicality plays a major role in access.

While hiragana mostly represent functional/grammatical morphemes and katakana loanwords, they overlap in actual usage in some interesting ways. The choice of script is not always consistent and may vary, depending on a writer's intention or even in-house publishing practices. For example, the word for 'eel' is pronounced /unagi/, but could be written in hiragana in one context, but as a katakana word in another context. When Kashu (1995) examined the subjective frequency of script type for 750 common words with Japanese college students, the results revealed that more than half of the words were identified as words which could be seen in more than one script.

Although individual preferences in the choice of script types are allowed, in general, a given kana word has a preferred kana type to be employed. There are then processing implications attached to bi-scriptalism. A common finding has been that words which are normally written in katakana are recognized faster in katakana than when they are written in hiragana. Conversely, words that are normally written in hiragana are recognized faster in hiragana than when they are written in katakana. For instance, based on reaction time differences in lexical access procedures for different types of kana strings, Sasanuma, Sakuma, and Tatsumi (1988) demonstrated that orthographically familiar kana words have a more direct access to the lexicon on the basis of the orthographic code, while orthographically unfamiliar words require recourse to phonological recoding. We may infer that visually familiar sequences of kana are treated as chunks in reading, in a way that
visually unfamiliar sequences are not (see also Kawakami, 1993), and that Japanese are definitely endowed with such script familiarity in processing of kana words.

Some work also indicate that the two syllabaries are domain-specific, with the main function of katakana tied to its representational function for foreign loanwords. Hatta, Katoh, and Kirsner (1984), for example, contrasted lexical decision rates for English-speaking learners of Japanese with native Japanese speakers, in order to tease out the nature of lexical representation for loanwords in Japanese. They suggest that native Japanese readers possess two separate, but partially overlapping lexicons: a foreign word lexicon to which katakana corresponds and a Japanese word lexicon to which hiragana script corresponds. In contrast, native English learners possess only one lexicon, a Japanese word lexicon, to which both hiragana and katakana correspond, suggesting that these learners lack script familiarity in processing given kana words.

5. Katakana Loanwords for Learners

As far as foreign learners of Japanese are concerned, katakana words seem to be strictly Japanese. Beginners, and even advanced learners, often complain that katakana words are difficult to decipher (see Kess & Miyamoto, 1999). We can think of a few reasons why processing such katakana borrowings from one's own language poses processing difficulty. First, there is the question of script familiarity cross-linguistically, as well as within Japanese itself. The switch from Roman alphabet to katakana removes the framework of graphic familiarity that easily identified the original borrowed form in its native English orthography. English-speaking learners have insufficient familiarity with katakana loanwords, so that the holistic cues which are gained from script familiarity and which allow at-a-glance recognition is absent. Hence, every katakana loanword is a deciphering exercise until script familiarity is acquired. Secondly, the auditory speech image for the loanword is considerably differ from its original. Thirdly, this loanword transfer problem is compounded by the fact that some words were borrowed through the auditory medium, while others through the visual medium, hindering any employment of orthographic transfer rules.

6. A Pilot Study

Given the above as background, we conduct a pilot study dealing basically with 'script type familiarity'. We examine cross-linguistic distances between the learners' auditory images of loanwords and katakana representations of these loanwords. To be more specific, we examine the relationship between learners' judgments of auditory similarity and their actual performance in identifying katakana words. If the source and borrowing are seen as the same, we would expect to have a high similarity rating and a high success rate in identifying katakana loanwords. If, however, there is a low correlation, the two lexical addresses bear little similarity to one another in the learners' mental dictionaries for Japanese and English.

Participants: Subjects were 29 English-speaking undergraduates taking Japanese courses at the University of Victoria. Their language abilities were evaluated at the intermediate level.
Questionnaires: Two questionnaires were employed. The first questionnaire consists of 60 high frequency loanwords taken from a daily newspaper. Participants were asked to identify each of the 60 Japanese words and to provide its corresponding English source word. The second questionnaire aimed at probing phonological similarity ratings. Participants were asked to rate the phonological similarity between katakana and corresponding English words in each of 60 pairs by using a six-point scale.

Results and Discussion: Two important considerations arose from our study. First, even though all the stimulus words were fairly common in occurrence, the participants' knowledge of these words was far from perfect. Despite the facts that ample time were assigned for their task and that participants knew that these stimuli were all directly from English, the results underscored their inability to come up with appropriate katakana transliteration rules. The number and type of invasion errors gave evidence of much guessing on their part. In many cases, their transliteration rules gave an approximation of phonological shape, leading the learners to mere educated guesses as to an identity of a lexical item. This was evident right across the array of words in the questionnaire. This fact must have arisen from a lack of script familiarity rather than from lack of word familiarity, given the fact that all the stimulus words are common in occurrence. Secondly, the correlation between success in word identification and the subjective judgments of phonological similarity was only .68, a moderate correlation at best. While the students could see the phonological similarities when the English and Japanese lexical items were placed together, they could not always see them, when a Japanese word was placed in isolation, and its English source was sought.

Conclusion: What our pilot study indicates is that we cannot support the common misconception among Japanese teachers of foreign students that katakana words will be mastered without difficulty since they are originally from English. Our study supports the opposite possibility; learners are not facilitated by katakana, and are unable to take advantage of the visual familiarity offered by katakana for loanwords. The pedagogical implication of our study is then that the principles of how katakana words are created to transliterate loanwords should be taught more carefully to foreign learners of Japanese. Transliteration rules need to be sufficiently explained and exemplified. Also, given the increasing inventory of loanwords in Japanese, this area would be an ideal candidate for computer-assisted learning exercises as the vehicle of pedagogical presentation for incremental learning protocols in the transliteration shifts and their apparent exceptions.
References


