Psycholinguistic Nature of the Japanese Orthography

Katsuo Tamaoka
Matsuyama University, Matsuyama, Japan

1 Historical Perspective of the Japanese Orthography
   1.1 Historical Perspective of Kanji Development
   1.2 Historical Perspective of Kana Development
   1.3 Historical Perspective of Japanese Script Reform
   1.4 Summary

2 Phonetic Perspective of the Japanese Orthography
   2.1 Structure of the Japanese Sound System
   2.2 Phonetic Nature of Kana Script
   2.3 Homophonic Words in the Japanese Language
   2.4 Summary

3 Visual Perspective of the Japanese Orthography
   3.1 Kanji Categorization
   3.2 Radicals and Phonetic Elements of Kanji
   3.3 Kanji Content Ratio in Japanese Sentences
   3.4 Sense-Determinative and Sense-Discriminative Functions
       of Kana and Kanji
   3.5 Summary

Abstract: Beginning with the adoption of the Chinese language and proceeding to modern script reforms, the Japanese ortho-

Correspondence concerning this article should be addressed to Dr. Katsuo Tamaoka, Matsuyama University, 4-2 Bunkyo-cho, Matsuyama. # 790 Japan.
graphy is reviewed briefly from an historical perspective. After the historical discussion, the two types of script, kana and kanji, are discussed from a phonetic perspective to clarify the relationship between the Japanese sound system and the Japanese orthography. In addition to the phonetic argument, from the visual perspective, the Japanese scripts of kana and kanji are discussed in terms of activation units of kanji, an adequate kanji content ratio in Japanese sentences, and the sense-determinative and sense-discriminative functions of kana and kanji.

1 Historical Perspective of the Japanese Orthography

Although the history of the Japanese language is usually classified into five periods (Kindaichi, 1987), the Japanese writing system has developed most dramatically from a psycholinguistic perspective within three different periods (Hadamitzky & Spahn 1981; Kindaich, 1979; Miller, 1967; Sato, Hachiya, Kato, Hida, Sato, Suzuki, & Maeda, 1978; Seeley, 1984a; Tsukishima, 1979). In the Nara period (710-794), Japan adopted the Chinese written language. Japanese utterances were basically represented via the classical Chinese language. In the Heian period (794-1192), phonetic symbols were derived from Chinese characters and developed to phonetically represent Japanese spoken language. After the Meiji Restoration in 1867, when Japan was established as a modern nation, the Japanese orthography was standardized by the central power of the Japanese government.

1.1 Historical Perspective of Kanji Development

Japanese first encountered the written Chinese language around the end of the fourth and the beginning of the fifth century (Hadamitzky & Spahn,
1981; Kabashima, 1987; Seeley, 1984a; Tsukishima, 1979; Miller, 1967). The Japanese language is different from the Chinese language, especially in terms of sound system and syntax order. The Japanese sound system does not have the four tones which exist in the standard (Mandarin) form of the Chinese language (Kaiho & Nomura, 1983; Leong, 1986). The difference in sound systems makes the Chinese language difficult to adapt for representing Japanese utterances. In addition, unlike Chinese, Japanese has a wide variety of inflectional categories which usually arrange sentences in the order of subject, object and then verb. Because of these major linguistic differences, the Chinese script was inadequate to represent the Japanese spoken language (Kindaiichi, 1981; Miller, 1967; Seeley, 1984a). In the Nara period (710-794), when the first Japanese written documents appeared, Chinese was used as a written language and Japanese was separately maintained as a spoken language. Kindaiichi (1981) describes this period of the Japanese language as *genbunnito*, denoting the co-existence of two different languages — one spoken and the other written.

Since Japanese imitated kanji (Chinese characters) sounds without tones, pronunciations depended on the period during which Japan was in contact with China. Different types of pronunciation of kanji were borrowed from China during various periods (Hirai, 1979; Miller, 1967; Sato et al., 1978). Systems were labelled according to the names of dynasties in political power during each period that Japan came into contact with China. Kanji pronunciations are classified into three types: namely, *go-on*, *kan-on*, and *to-on* (formerly called *so-on*). These three different systems of pronunciation simultaneously exist in the pronunciation of kanji used in modern Japanese.

The earliest written text in Japan is *Kojiki* [*Records of Ancient
Matters], completed in 712. This text was an early Japanese historical document that recorded ancient events as recited by a court official. Except for songs and proper nouns (e.g., names of gods, Japanese persons and places in Japan) that were phonetically written in kanji, this document was written in the Chinese language commonly known in Japan as kanbun, the classical Chinese writing style (Kabashima, 1987; Kitayama, 1967; Tsukishima, 1979). In Kojiki, kanji is meant to be read according to the system of the Six Dynasties known as go-on. It was used in Buddhist terminology throughout the eighth century (Miller, 1967). Today, go-on continues to exist in words of modern Japanese pertaining to Buddhist terms.

In 720, a more sophisticated historical text, Nihon Shoki [The History of Japan], was completed. This document was intended to be read according to the kan-on, as used in Ch’ang-an at the peak of the T’ang Dynasty (Miller, 1967). Just as go-on had won its place in the area of Buddhism, kan-on became accepted in the area of Confucianism and other non-Buddhist Chinese studies. Later in the Heian period (794-1192), kan-on was officially regarded as the standard pronunciation in Japan (Sato et al., 1978; Hirai, 1975). “Many common Chinese loanwords in Japanese today are old blends of the two systems” (Miller, 1967, p. 107).

Much later, in the fourteenth century, to-on (the third sound system of kanbun) was brought to Japan, mostly by Zen Buddhists who had studied in China. To-on, an ancestor of modern Pekingese (Miller, 1967), was the common language of north China at the time. Novel things brought back by Zen monks (e.g., andon [paper-covered night light], chochin [lantern], and noren [shop curtain]), as well as Zen Buddhist terms, were named according to to-on and these words remain in the modern Japanese lan-
guage (Sato et al., 1978; Hirai, 1979).

Japan's deep historical involvement with China, which extended from the period of the Six Dynasties to the T’ang and Sung Dynasties, contributed to enlarging the Japanese vocabulary with these words pertaining to Buddhism, Confucianism and Zen. Since the sound system of the Japanese language does not have the four tones of the Chinese language, adapting kanji pronunciations from China resulted in the co-existence of great numbers of homophonic kanji. Multiple pronunciations of single kanji not only made it difficult to associate the sound with meaning, but also often created various multi-lexical kanji. For example, a kanji, 行 has three different pronunciations and meanings based on the period from which it was adapted: 修行 [shugyo], meaning training, from go-on; 孝行 [kou-kou], meaning dutifulness, from kan-on; and 行脚 [angya], meaning pilgrimage, from to-on. It is often difficult to find a proper kanji pronunciation since a single kanji could be read differently depending on its usage in combination with other kanji.

1.2 Historical Perspective of Kana Development

From the Nara period (710-794) to the end of the Edo period (1868), kanbun (the classical Chinese writing style) was used as the official written language, just as Latin was used in scholarly writings in medieval Europe. Because of the difficulty involved, until the Heian period (794-1192) Japanese seemed to have no intention of reforming kanbun to present spoken Japanese (Kabashima, 1987).

The need to transcribe Japanese names and poems was satisfied by using the kanji not for their meanings, but for their sounds (Seeley, 1984a). Around 760, Manyoshu, a collection of approximately forty-five hundred
poems, was written phonetically in kanji. In this anthology of poetry, almost all the poems were written using the syllable and kanji correspondence technique known as *man’yogana* (Tsukishima, 1979; Sato et al., 1978), with some exceptions originally written in the *kanbun* style (Kabashima, 1987). However, since *man’yogana* utilized kanji without limiting the number of kanji characters and without simplifying them, it proved to be an overly cumbersome method of representing Japanese sounds (Miller, 1967).

Although the linguistic technique of *man’yogana* was impractical for transcribing the Japanese language, it became an early step toward devising a simpler form of representing Japanese syllables, known as kana. Based on *man’yogana*, two types of kana were formed in the Heian period by abbreviating kanji in two ways. One technique involved cursivizing and the other involved isolating (Seeley, 1984a). The cursivizing technique represents Japanese sounds by simplifying the figures of kanji, thus forming hiragana. Katakana, on the other hand, was created by the isolating technique, which represents sounds by extracting a simple part of the more complicated kanji shape. In the historical process of developing kana, hiragana and katakana were created independently. Hiragana was primarily developed through women’s literature for representing Japanese utterances, whereas katakana was developed by Buddhist priests as an aid for reading *kanbun*.

About 1002, *Genji Monogatari* [*Tale of Genji*], consisting of fifty-four volumes, was written in hiragana — with a few kanji words — by a daughter of Fujiwara Tametoki known as Murasaki Shikibu (Miller, 1967; Sato et al., 1978; Tsukishima, 1979). She depicted the court life of the hero Genji, a prince by birth, who was surrounded by many talented women
but failed to become an emperor. Miller (1967) describes Japanese writing in those days as follows:

...although the language for official documents and other important uses of writing remained Chinese, novels, diaries, and other frivolous works, especially works by and designed largely for women, were generally written in Japanese (p. 39).

Since hiragana was developed to present sounds of the Japanese language in the sequence of Japanese syntax order, hiragana was very often used to record stories and daily events, especially by women. Hence, early written Japanese was mostly developed by intelligent women of the nobility.

In the early Heian period, small numbers of Japanese intellectuals began to accompany the Chinese rendering with a Japanese translation. Japanese Buddhist priests started to add katakana as a phonogram to help them remember the pronunciation of the kanji used in Chinese texts. Later, these Chinese texts were translated by rearranging the order of characters and providing additional elements of katakana required by the conventions of Japanese syntax (Seeley, 1984a). At this time, katakana came into general use, (apart from its continuous popular use as a mnemonic and additional element of Japanese inflection) in reading and rendering Chinese text. Later both hiragana and katakana were used to denote inflectional categories of Japanese syntax. Thus, this phonological syllable-kana correspondence came to represent both native Japanese sounds and inflectional categories.
1.3 Historical Perspective of Japanese Script Reform

Language education in the Edo period (1603-1867) emphasized kanji writing and reading in order to refine the general knowledge of the samurai (soldier) and upper-business classes (Mabuchi, 1983). In 1868, the year of the Meiji Restoration, the Japanese language was not an effective means of communication for the general public. Twine (1983) summarized the problems of the Japanese language in those days:

There was a complicated network of regional dialects; the spoken and written languages were so dissimilar as to necessitate the complication of separate grammars for each; the written language itself was divided into several discrete styles, each drawing its vocabulary and syntax from early medieval Chinese or Japanese; and there were more than ten thousand Chinese characters in use (p. 115).

Thus, after the Meiji Restoration, the major task of government language reform was to simplify the Japanese orthography for effective use in daily communication by the general public.

The major proposals for reforming the Japanese language fell into three categories: (1) replacement of kanji by kana; (2) use of romanization; and (3) continued use of kana and kanji, with restrictions on the number of kanji (Seeley, 1984b). In 1866, just two years before the Meiji Restoration, an early proposal of script reform was presented by Hisoka Maejima to the Shogun, suggesting the complete replacement of kanji by kana in order to remove the great barrier of writing and reading kanji for the general public (Twine, 1983). Maejima’s proposal later led to the establishment of the Kana Club in 1882 and the Roma-ji Club in 1885, which
intended to change every kanji character to the phonetic scripts of kana or roma-ji (romanization), respectively. However, Maejima and his supporters’ intention to democratize the writing system was too drastic and radical to convince the bureaucracy.

Since 1900, the Ministry of Education has been reforming Japanese script to adjust the existing writing system to suit the general public. A series of major reforms by the government included: (1) limiting the number of kanji characters in use and (2) standardizing kana and its usage. Seeley (1984b) has concisely explained the script reforms undertaken by the Japanese government since 1900.

In 1900, the Ministry of Education set the first regulations on standard kana signs and kana usage, and restricted the number of kanji taught in primary school to about 1,200. These regulations were rescinded, however, in 1908. In 1923, Rinji Kokugo Chosakai (Interim Committee on the National Language) produced *Jouyo Kanjihyo* (a list of kanji for general use), consisting of 2,108 kanji and limiting the use of kanji outside of the list other than in proper nouns, the Imperial rescripts, legal texts and unavoidable quotations. The *Jouyo Kanjihyo* was revised and publically proposed in 1931 as a list of 1,856 kanji. However, Seeley (1984b) explained that the Manchurian Incident in 1931, which created a mood of conservatism and nationalism, produced an obstacle to revising kanji usage.

After the defeat of Japan in the Second World War, the Allied Force High Command requested the Ministry of Education to limit the number of kanji used in textbooks. After several sessions at the government level, the *Jouyo Kanjihyo* (a list of kanji for current use) was created in 1946 (Seeley, 1984b). This list, consisting of 1,845 kanji, became the standard of kanji learning for compulsory education from Grades 1 to 9. In the same
year, the Ministry of Education introduced *Jouyo Kanji Beppyo* (a separate list of Jouyo Kanji) for educational purposes for Grades 1 to 6 (Kadokawa Publishing Company, 1978).

In 1971, with the addition of 115 kanji to the *Jouyo Kanji Beppyo* (881 kanji), the number of kanji taught in Grades 1 to 6 was specified: 76 kanji for Grade 1, 145 for Grade 2, 195 for Grade 3, 195 for Grade 4, 195 for Grade 5, and 190 for Grade 6 (the Ministry of Education, 1987). The remainder were taught at the junior high school level (Grades 7–9), so that all the 1,850 kanji listed in the *Jouyo Kanjihyo* should be taught by the end of Grade 9 (The Ministry of Education, 1986). The standardization of kanji education has set the clear goal of accomplishing kanji learning during compulsory education (Grades 1 to 9).

The kanji ratio in terms of the total kana and kanji contained in sentences in Japanese newspaper articles showed a dynamic change through the process of kanji standardization. In 1879, the kanji ratio in Japanese newspaper articles was about 94 percent, whereas in 1968 it became about 75 percent (Hayashi, 1982). In contrast, limiting kanji use created a difficulty in writing newly-adapted words from alphabetic languages. Especially after the Meiji Restoration, various words from Europe have flowed into the Japanese language. Before the Second World War, adaptation of an alphabetic loanword was mostly done by describing its meaning by kanji combinations. However, because the kanji limit for daily use was set by the Japanese government, which also narrowed the range of kanji expression, katakana came to be the most efficient and popular way of representing alphabetic loanwords.

Until recently, katakana and hiragana usually appeared in separate texts. In modern Japanese writing, after the Second World War, the three
types of Japanese scripts are used for different purposes within the same text: (1) kanji has continued to represent Chinese loanwords within the outline of the standardization; (2) hiragana has continued to describe Japanese inflectional categories; and (3) katakana is now used for presenting sounds of loanwords from alphabetic languages such as Spanish, Dutch, English, German and French. Consequently, in the modern Japanese language, kanji and hiragana continue their original roles, while katakana has changed its purpose from a phonetic aid for reading Chinese texts (or their translations) to depicting sounds of loanwords from foreign alphabetic languages.

1.4 Summary

In this section, the origin and development of Japanese scripts were discussed from a historical perspective. Kanji was adapted from Chinese characters and three different sound systems, resulting in various homophonic kanji pronunciations. Hiragana and katakana were developed domestically to represent spoken Japanese. In the modern Japanese language, kanji and these two types of kana are used in the same sentence. Kanji represents various words, mostly nouns and adjectives. The two types of kana serve different purposes in expressing Japanese utterances: hiragana is used mainly for inflections and adverbs; and katakana for alphabetic loanwords.

Unlike the English and Chinese languages, the Japanese language contains three different scripts, varying from morphemic kanji to phonetic kana. There must be a reason for the co-existence of these three scripts in the Japanese language apart from the historical perspective. Therefore, the next section discusses characteristics of the three different scripts from
a phonetic perspective in order to clarify the phonetic characteristics of the Japanese orthography.

2 Phonetic Perspective of the Japanese Orthography

From the phonetic perspective, the Japanese sound system is discussed in terms of its phonemic procedure for constructing syllables. Following the analysis of the Japanese phonetic structure, two types of kana, hiragana and katakana, are described on the basis of their phonemic combinations and their representation of Japanese utterances. Finally, various homophonic words existing in the Japanese language are discussed as one of its major phonetic characteristics.

2.1 Structure of the Japanese Sound System

Compared to English, the Japanese language has a simple vowel system. Maddieson (1984) classified the Japanese language as a 5-vowel system, whereas English contains 12 vowels. Despite this small number of vowel sounds, as with most of the Austronesian group, the Japanese language could be characterized as a vowel-oriented sound system because of high vowel-to-consonant ratios. For example, there are many frequently-used words composed only of vowel sounds such as: ai, meaning love; au, meaning meet; ue, meaning above; ao, meaning blue; and ie, meaning house. Among Japanese vowel sounds (i.e., /a/, /i/, /u/, /e/ and /o/), the /a/ sound is the most frequently used and the /e/ sound is the least frequent (Kindaichi, 1979). Thus, the /a/ vowel sound represents the most common sound in the Japanese language.

English has 24 consonants (Crystal, 1987), whereas there are only 14
consonants and two semivowels /w/ and /y/ in Japanese (Kindaichi, 1979; Mizutani, 1981). Since Japanese syllables are required to end with vowel sounds (except the four geminated consonants /p/, /k/, /s/ and /t/), the possible consonant and vowel combinations of syllables used in the Japanese language are very limited, comprising only 101 combinations. The sound /n/ is counted as one since /n/ is a morphophonemic syllable for the alternation of nasals /m, n, ng/ or nasalization of the preceding vowel when not followed by a homorganic consonant. This is very different from the construction of English syllables, where there are 429 syllables ending with /d/, 383 ending with /z/, and 376 ending with /t/ (Rockey, 1973). The number of syllables ending with the ten most common consonants adds up to 2,963 possible combinations. In the Chinese sound system the four tones, which are the pitch of the voiced part of the character, create a great sound variation and play an important role in distinguishing homophonic characters (Leong, 1986). In contrast to the great variety of sounds existing in English and Chinese, the Japanese language is confined to a limited combination of 101 syllables. However, there are 104 kana symbols for 101 syllables, since three kana symbols overlap their syllabic sounds. From the phonetic perspective, this small number of syllables could be the major linguistic characteristic of the Japanese sound system.

As Japanese syllables are pronounced with approximately equal duration for each syllable, the syllable in the Japanese language is technically called mora (plural is morae), which is derived from a Latin word meaning delay or space of time (Bunkacho, 1979; Iritani, 1983). For instance, *watashi*, meaning "I", is pronounced in three syllables /wa/, /ta/, and /shi/ with an equal length of time duration. The mora is the smallest sound unit in the Japanese language; therefore, the mora may be the
smallest unit of phonological coding used to access the Japanese mental lexicon. Unlike the English language, morae do not contain the combination of consonant + vowel + consonant, except /k/, /p/, /s/ and /t/, which can appear between two morae as a dual consonant when constructing words. This is another reason that only 101 syllables or mora combinations exist in the Japanese language.

The constructions of mora are categorized into five possible combinations: (1) single vowel; (2) consonant + vowel; (3) semivowel + vowel; (4) consonant + semivowel + vowel; and (5) special sounds (Mizutani, 1981; Satou et al., 1978; Tsukishima, 1979). A mora can be constructed

<table>
<thead>
<tr>
<th>Type of Mora Combination</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combination #1: Single Vowel</td>
<td>5</td>
</tr>
<tr>
<td>Combination #2: Consonant + Vowel</td>
<td>58</td>
</tr>
<tr>
<td>Combination #3: Semivowel + Vowel</td>
<td>4</td>
</tr>
<tr>
<td>Combination #4: Consonant + Semivowel + Vowel</td>
<td>33</td>
</tr>
<tr>
<td>Combination #5: Special/n/</td>
<td>1</td>
</tr>
<tr>
<td>Total Combinations</td>
<td>101</td>
</tr>
</tbody>
</table>

*Note:* /shi/, /chi/, /tsu/, /sha/, /shu/, /sho/, /cha/, /chu/, /cho/, /ja/, /ju/, and /jo/ are classified as the second type of combination. /k/, /p/, /s/ and /t/ also appear at the end of mora as a double consonant which could be categorized as the fifth type of combination. Since these consonant sounds occur only as a phonetic change under certain conditions, they are excluded from the table.

with a single vowel sound, but since there are only five vowel sounds in the Japanese language, the five combinations exist as described in Table 1. The second combination of consonant + vowel is the fundamental feature of morae, having 58 mora combinations. The third type, semivowel + vowel combinations, has four different kinds: /ya/, /yu/, /yo/, and /wa/. The fourth type, consonant + semivowel + vowel, is the second most
common combination and has 33 different mora combinations. Finally, there are some exceptional consonants: /n/, counted as one syllable, can appear at the end of words, and /k/, /p/, /t/ and /s/ can be placed as a dual consonant between syllables such as /ka/ + /pa/ = /kappa/, /ka/ + /ta/ = /katta/, and /ka/ + /ka/ = /kakka/ (Mizutani, 1981).

Since Japanese syllables must end with a vowel sound, only five kinds of combinations can possibly be used to construct syllables. Possible syllable constructions using phonemes are limited to 101 different types. In the Japanese orthography, each kana represents exactly one syllable or mora (with three exceptions /o/, /ji/ and /zu/). In the next section, kana, the phonetic script of the Japanese orthography, is discussed in order to examine the relationship of phonemic sounds to kana script.

2.2 Phonetic Nature of Kana Script

Kana phonetically differs from the English alphabet in two fundamental ways (Kimura, 1984). First, kana symbols represents syllables of spoken Japanese, while English symbol-and-sound correspondence takes place at the level of phonemes. Second, kana has a one-to-one correspondence between symbol and sound, whereas the English alphabet involves many irregularities, making it more difficult to apply a phonological strategy to lexical access.

There are two kinds of kana — hiragana and katakana. In modern Japanese, kana includes 46 basic syllables. The chart is called the standard “Chart of the 50 Sounds”. After the kana syllabaries had been constructed, the consonant /n/ was added to the chart (Hadamitzky & Spahn, 1981) and the chart included the 46 syllables.

Although the basic sounds in the Japanese language are constructed by
combinations of the five vowels with the seven consonants and two semi-
vowels, there are some exceptions: /shi/ in the /s/ sound column, /chi/ and /tsu/ in the /t/ sound column, and /fu/ in the /h/ sound column. In addition, the syllables are expanded to 71 kana by adding 25 more syllables through the use of diacritics, symbolized by two small dots or a small circle. With the semivowel /y/ and 11 consonants, these syllabaries are further expanded into 104 kana and presents the basic sound structure of the Japanese language. Among the 104 kana, three actually represent the same mora sounds: お and を for /o/; じ and ぢ for /ji/; and づ and ず for /zu/. However, these are customarily used in the context of describing different words. Hence, the 104 kana, including kana expanded by two kinds of diacritics, represent the 101 syllables existing in the Japanese language. The two types of kana described are used in different contexts, hiragana for inflection in the Japanese syntax and katakana for representing words and names borrowed from alphabetic languages.

According to the study by the Japanese National Language Research Institute (Kokuritu Kokugo Kenkyujo, 1964), only 36.7 percent of the Japanese vocabulary used in 90 magazines published between 1950 and 1956 was indigenous. 47.5 percent was of Chinese origin, 9.8 percent originated from alphabetic languages and 6.0 percent was comprised of hybrids. Within the 9.8 percent of the vocabulary which originated from alphabetic languages, loanwords from English constituted 80.8 percent, or 7.92 percent of the total Japanese vocabulary. Loanwords of alphabetic origin are printed in katakana script, so they are visually distinct from words printed in hiragana and kanji.

The small number of syllables also limits the number or words phonologically constructed by syllables. This linguistic nature of Japanese
kana has created various homophonic words in the Japanese vocabulary. Because of the limited number of syllables in the Japanese sound system, visual differentiation of the three scripts could be fundamental for accessing the Japanese mental lexicon. In daily life, Japanese can be seen tracing different kanji on the palms of their hands to identify a correct kanji from many homophonic words. In order to distinguish efficiently various homophonic words, three different scripts are necessary. In the next section, homophonic words in the Japanese language are discussed to examine the need for the three scripts in writing the Japanese language.

2.3 Homophonic Words in the Japanese Language

Kanji reading can be divided into two types: the on-reading derived from the original Chinese pronunciation, and the kun-reading originating from the Japanese way of reading kanji. When Japanese borrowed Chinese characters, they ignored the four tones in the Chinese pronunciation and adapted the three different kanji on-readings called go-on, kan-on and tou-on. In simplifying pronunciations of the Chinese tones and adapting three different sound systems from China, the Japanese created a great number of homophones in kanji. In addition, each kanji usually involves one or more Japanese original readings, called kun-readings. This linguistic nature of the Japanese kanji reading system created great difficulties in accessing the mental lexicon.

Morton and Sasanuma (1984) suggested that the real complexity in identifying the pronunciations of compound kanji, comprised of single kanji, occurs because there are no systematic rules for determining the on-reading or kun-reading of a single kanji. According to Kaiho and Nomura (1983), as an extreme case, two kanji characters out of the 1945
Jouyo Kanji possess two kinds of on-reading, and 10 kinds of kun-reading (see Table 2). For the on-reading, 145 kanji (7.46%) out of the 1945 Jouyo Kanji have more than two kinds of on-reading, and only one kanji has three on-readings. For the kun-reading, 278 kanji (14.29%) have two kinds, 91 kanji (4.68%) have three kinds, 40 kanji (2.06%) have four kinds, 7 kanji (0.36%) have five kinds, and 7 kanji have six or more kinds. For example, a commonly used kanji 日, meaning a day, the sun or sunlight, has two on-reading, /nichi/ and /jitsu/, and two kun-reading, /hi/ and /ka/. Although the majority of kanji have one on-reading and one or no kun-

<table>
<thead>
<tr>
<th>Number of Kun-Readings</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>694</td>
<td>43</td>
<td>-</td>
<td>-</td>
<td>37.89</td>
<td>737</td>
</tr>
<tr>
<td>1</td>
<td>32</td>
<td>691</td>
<td>61</td>
<td>-</td>
<td>40.03</td>
<td>784</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>247</td>
<td>24</td>
<td>-</td>
<td>14.29</td>
<td>278</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>82</td>
<td>8</td>
<td>-</td>
<td>4.68</td>
<td>91</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>33</td>
<td>6</td>
<td>1</td>
<td>2.06</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>7</td>
<td>-</td>
<td>-</td>
<td>.36</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>.05</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>.21</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.00</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>.05</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>.10</td>
<td>2</td>
</tr>
<tr>
<td>%</td>
<td>2.06</td>
<td>90.44</td>
<td>7.46</td>
<td>.01</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>1,759</td>
<td>145</td>
<td>1</td>
<td></td>
<td>1,945</td>
</tr>
</tbody>
</table>

Note: The special on-readings and kun-readings are excluded from the table. The percentages are added by the author. (Kaiho & Nomura, 1983, p. 127)
reading, it is usually impossible to identify the single proper pronunciation for a single kanji without referring to the context.

According to the Japanese word-processor Sharp WD-55B, a given syllable like /a/ can have as many as five different writings in kanji (e.g., 哉, 阿, 娃, 蛙), 34 kanji for a sound /i/ (e.g., 胃, 伊, 意, 位, 委), 12 kanji for a sound /u/ (e.g., 右, 汪, 字, 鳥, 羽), 17 kanji for a sound /e/ (e.g., 絵, 柄, 江, 恵, 重), and 8 kanji for a sound /o/ (e.g., 緒, 御, 悪, 於, 汚). Furthermore, Iwabuchi (1987) found in the Japanese dictionary 61 ways to write the sound /kou/ in kanji, 63 ways for the sound /shou/. 47 ways for /shi/, and 44 ways for /kan/. Again, the sound itself cannot identify a proper kanji without referring to the context.

Japanese nouns are usually constructed by combining two kanji. Homophonics of two kanji combinations also present the same problems as single kanji (Iwabuchi, 1987). There are 27 ways of writing the sound /kikou/ (e.g., 気候, 機構, 紀行), and 29 ways for the sound /koushou/ (e.g., 交渉, 校長, 口承). Each of these kanji combinations has a different meaning, but the proper combination of kanji can usually be found from the context. Some nouns composed of two kanji, however, have similar meanings. For example, /kagaku/ could mean ‘science’ (科学) or ‘chemistry’ (化学), although ‘chemistry’ is often pronounced /bakegaku/, indicating its meaning as the science of changing molecules. Nevertheless, without referring to the context, proper kanji combinations for a particular noun, as with single kanji, still cannot be identified merely by having just a pronunciation of the word.

Japanese children learn nouns constructed with compound characters through writing the phonetic kana symbols above or to the side of the kanji. An children memorize various compound kanji forming mostly nouns and
verbs, these kana aids are gradually removed from the school textbooks. Using this learning procedure of Japanese children as evidence, Saito, Inoue and Nomura (1979) proposed that kanji characters can be processed phonologically. However, phonological codes of kanji become unreliable, since kanji pronunciations overlap each other. Therefore, the memory storage used in reading kanji would be filled with multiple pronunciations for some kanji and their possible combinations for constructing words.

Homophonic words are also found in words in katakana (Iwabuchi, 1987). Because there is no sound distinction in the Japanese language between /r/ and /l/, 'grammar' and 'glamour' are transcribed in katakana in the same way: グラマー, pronounced /guramaa/. The sounds /v/ and /b/ also cannot be distinguished clearly in the Japanese sound system, so that 'valve' and 'bulb' are described in the same way: バルブ, pronounced /barubu/. The equal duration of syllables required in the Japanese sound system make original loanwords difficult to guess: 'morning' is pronounced /mouningu/, 'mass-communication' becomes the very complex of /masukomyunikeishon/. Large numbers of homophonic words written in katakana were rarely seen in the Japanese vocabulary until recently, since the number of loanwords from alphabetic languages are relatively small compared to Chinese loanwords.

Japanese homophonic words are sometimes distinguished by an accent. There are two types of accent; the pitch accent used in the Chinese four tones and the stress accent used in English. In the Japanese language, the pitch accent is occasionally used to differentiate homophonic words (Tsukishima, 1979; Satou et al., 1978). For example, the sound /hashi/ signifies three commonly-used words — chopsticks, bridge, and edge. According to the standard Tokyo accent, /hashi/ for 'chopsticks' is pronounced with
a high pitch for /ha/ and a low pitch for /shi/. For 'bridge', this pattern is reversed to a low pitch for /ha/ and a high pitch for /shi/. For 'edge', both syllables are pronounced with the same pitch (Iwabuchi, 1987). The practice of pitch accent, however, differs depending on dialect. For example, the Osaka dialect often utilizes an opposite pitch pattern from the standard Tokyo accent. The height of a sound is therefore not an efficient way of distinguishing homophonic words. This may be one of the reasons that kanji remained a major part of the Japanese orthography as a visual distinction for homophonic words, although there were several simplification movements that proposed completely abolishing kanji from the Japanese language.

2.4 Summary

Historically, the Japanese adapted Chinese characters without the four tones, which resulted in multiple pronunciations. Phonetically, Japanese syllables are limited by the five patterns of phonemic combinations which construct 101 syllables. This limited number of syllables or morae has created two phonetic characteristics of kanji: (1) multiple pronunciation of a particular kanji and (2) multiple kanji for a particular pronunciation. Kanji's continued existence in the Japanese language as a major orthography could be attributed to this distinctive characteristic of the simple Japanese sound system. The co-existence of kanji and two different kana helps in visually distinguishing miscellaneous words in a sentence; furthermore, kanji plays an important role in visually distinguishing homophonic words. The next section discusses the visual perspective of the Japanese orthography, focusing on kanji in particular.
3 Visual Perspective of the Japanese Orthography

Since the Japanese sound system is limited to 101 syllables in the standard kana phonetic chart, phonetic recording might be too restricted in accessing the Japanese mental lexicon. In this section, the major concern shifts to the visual perspective, especially of kanji. First, the categorization of kanji is discussed in order to calculate the ratio of phonetic kanji in the 996 basic kanji. Second, the radicals used in constructing kanji are examined to clarify the activating element of lexical access. This accompanied by a discussion of kanji phonetic elements. Third, the kanji content ratio in sentences is reviewed to understand the visual distinction from kana. Finally, two basic functions of visual orthographic recognition — sense-determinative and sense-discriminative functions — are discussed in the relationship between kanji complexity and recognition time.

3.1 Kanji Categorization

The most common categorization of kanji is the Chinese device of six categories of characters — which Japanese also adapted for classifying kanji (Hirai, 1979; Leong, 1986; Tsukushima, 1979; Satou et al., 1978). These six categories are: (1) ‘pictographic kanji’ derived from imitating the shapes of objects; (2) ‘ideographic kanji’ which indicate ideas and qualities; (3) ‘compound ideographic kanji’ formed by combining more than two component kanji parts to represent ideas and their associations; (4) ‘phonetic compound kanji’ constructed by the phonetic and semantic components; (5) ‘loan kanji’ adapted for recording new kanji by sound, independent of meaning; and (6) ‘analogous kanji’, which are new kanji patterned
after old kanji to denote new meanings.

According to these six categories of kanji, Watanabe (1976) classified 996 kanji (Gakushu Kanji [A List of Learning Kanji]) which are taught from Grades 1 to 6 in Japan. Among the 996 kanji, ‘phonetic compound kanji’ comprised 45.6 percent, whereas ‘pictographic kanji’ made up only 11.5 percent. Ito (1979), likewise, counted the number of kanji according to the four categorizations of ‘pictographs’, ‘ideographs’, ‘compound ideographs’ and ‘phonetic compounds’. Out of 996 taught by the end of Grade 6, 553 kanji or 55.5 percent are classified as ‘phonetic compounds’. Out of 1,850 Jouyo Kanji taught by the end of Grade 9, 1,211 kanji or 65.5 percent are classified as ‘phonetic compounds’. Ito used 1,933 kanji altogether for her study, and found that 1,278 kanji or 66.1 percent of the total are categorized as ‘phonetic compounds’.

Kanji is often referred to by many psycholinguists and even some linguists as a logographic or pictographic script (Bryant & Bradley, 1983; Gelb, 1964; Henderson, 1982; Kimura, 1984; Makita, 1968; Morton and Sasanuma, 1984; Read, 1983; Sasanuma, 1977). Pictographic kanji, which are often quoted as typical examples of kanji’s development from the shape of objects, actually comprise only about 10 percent of the 996 basic kanji. Therefore, although the pronunciation of one kanji frequently overlaps with those of other kanji, phonetic elements of kanji may well be used for lexical access along with its pictographic or semantic elements.

3.2 Radicals and Phonetic Elements of Kanji

Based on the number of kanji combinations, Morton and Sasanuma (1984) classified kanji into two types: single characters and compound characters (consisting of a sequence of single characters). The single
characters are further divided into two types, simple and complex. There are 214 fundamental components or ‘radicals’, including all the single characters that make an independent kanji meaningful.

The characters in the Chinese language also contain 214 radicals, which were identical to those in Japanese until the Chinese government simplified the writing system. In a study of 8,711 Chinese characters, 17 out of 214 radicals were utilized as an element in constructing 50.17 percent of the characters (Leong, 1973). Since the traditional Chinese orthography has shown remarkable stability regardless of the years and geographical distance (Leong, 1986), radicals used in Japanese kanji should show results similar to those indicated in the study of Chinese characters. Accordingly, commonly-used radicals in Japanese kanji are limited to a small number.

Combinations of two or three radicals make up complex characters. However, since pictographic kanji characters are a minority, these combinations rarely represent the meanings of the complex characters. In addition, the kanji characters reflect a type of culturally shared knowledge which is not as immediately available for the contemporary speakers of the Japanese language. For instance, the tree radical found in the kanji character, /hashi/ meaning ‘bridge’, may have been an appropriate meaning for an age when bridges were built of wood. In the modern kanji construction, the character of ‘bridge’ should employ an ‘iron’ radical instead.

Furthermore, combining multiple radicals or kanji elements often results more confusing as to the specific meaning. For example, the kanji character 注 /chuu/ meaning ‘notes’ or ‘interpretation’ is constructed by two parts, a radical .GetHashCode() indicating ‘water’ and an element 主 meaning ‘pole’. This character represents the meaning of water pouring straight. Again, complex kanji characters in the modern Japanese language are very often
too abstract to determine from its radicals and elements.

Kanji are often constructed using other components in combination with the 214 radicals. The number of phonetic elements of kanji was calculated by Ito (1979). She counted the number of phonetic kanji and found 1,278 phonetic compounds out of 1,933. However, among 1,278 phonetic compounds, 30 kanji are *kun*-reading—the Japanese original way of reading—and 56 kanji are abridged or have lost their original sounds. Consequently, Ito suggested that 1,192 kanji \((1,278 - 30 - 56 = 1,192)\) could be genuinely counted as ‘phonetic compounds’. But this figure still represents 61.7 percent of the total 1,933 kanji. Among the 1,192 kanji genuinely classified as ‘phonetic compounds’, Ito found 633 different kinds of phonetic elements, further indicating that the most frequently used phonetic radical is 方 [hou], forming nine different kanji. According to her analysis, Ito (1979) suggested that:

... in the process of teaching, if kanji were divided into phonetic and semantic elements, it would be mastered more efficiently knowing these phonetic building blocks (p. 75).

The radicals can be classified from their shapes into seven categories; namely, *hen* (side), *tsukuri* (building), *kanmuri* (crown), *ashi* (leg), *kamae* (structure), *tare* (hanging), and *nyo* (entering) in Japanese terms. These categories can be further arranged spatially as drawn in Figure 1. Since these seven categories give the basic spatial arrangement of radicals constructing a majority of kanji, a specific radical used in a target kanji can be found according to the recognition of the spatial arrangement of radicals. Then, the target kanji can be visually processed to find phonetic and
semantic elements used for deriving its meaning. Thus, the basic shapes of radicals can be one of the activating elements in processing kanji for lexical access.

3.3 Kanji Content Ratio in Japanese Sentences

The number of kanji required to read Japanese newspapers and magazines is often estimated to be about 3,000 (Sasanuma, et al., 1977). In fact, 3,000 kanji covers about 99.9 percent of the kanji used in newspapers and magazines (Kobayashi, 1981). Moreover, Kobayashi indicated that the 1,000 most frequently used kanji represent 93.9 percent of the kanji printed
in newspapers, and 90.0 percent of the kanji appearing in magazines. The knowledge of 2,000 kanji covers 99.6 percent of the kanji in newspapers and 98.6 percent of the kanji in magazines. Thus, the difference between 2,000 and 3,000 kanji characters could yield a margin of 0.3 percent of the total kanji in newspapers and 1.3 percent of those in magazines.

The average ratio of kanji in the total of kana and kanji contained in Japanese writings is estimated at approximately 30 percent (Kaiho & Nomura, 1983). The highest ratio, 42 percent was found in newspapers; 33 percent was found in 100 examples of modern literature; 25 percent appeared in a series of books published by Iwanami Publishing Company; and 30 percent in 27 different magazines. The high ratio of kanji contained in newspapers could be attributed to frequent references to proper nouns and places, which are usually printed in kanji. Among the writings of university students, a ratio of 30 percent of kanji was found in essays and 32 percent in critiques (Kaiho & Nomura, 1983). The overall approximate ratio of 30 percent may indicate an optimal incidence of kanji in Japanese sentences. Kanji can be easily recognized as figures in relief against the background of kana, enhancing reading efficiency.

Although mature Japanese readers have a knowledge of the 3,000 kanji which encompass 99.9 percent of kanji printed in Japanese newspapers and magazines, it is unrealistic for a reader to have to search all 3,000 kanji in their memory whenever they see a new kanji in context. Because of the inefficiency of activating all these kanji together to search for the correct meaning of a single target kanji, the major radicals may act as an important key for activating a specific block of kanji out of the entire kanji memory storage. Parallel to semantic elements of kanji, phonetic elements also help in activating the meaning of a target kanji.
3.4 Sense-Determinative and Sense-Discriminative Functions of Kana and Kanji

The sense-determinative function and sense-discriminative function (Kaiho, 1987; Leong, 1987) of the orthography will play an important role in decoding kanji and kana figures to access the lexicon. In the context of the Japanese orthography, the sense-determinative function refers to a script function where each kanji or kana is identified by its shape and sound, whereas the sense-discriminative function implies a script function where each kanji or kana is distinguished from others. These two functions are especially critical elements for activating target kanji and kana from the mental dictionary.

One of the activation elements of kanji and kana might be related to the number of strokes which are required to write them. The average number of strokes for writing the 46 basic hiragana is 2.3 strokes with a standard deviation of 0.7; for the 46 basic katakana it is also 2.3 strokes, but its standard deviation is slightly higher at 0.9. To complete the 26 capital letters of the English alphabet, an average of 2.0 strokes with 0.9 standard deviation are required (Kaiho & Nomura, 1983). As such, hiragana and katakana require a similar average number of strokes as the capital letters of the English alphabet.

The average number of strokes for the 1,945 多用 Kanji characters was much higher than for both kana, at an average of 10.3 strokes with a standard deviation of 3.8 (Kaiho, 1987). Leong (1973) analyzed the commonly used 1,851 Chinese characters and found an average of 11.27 strokes with a standard deviation of 4.45. The higher number of strokes and the wider deviation reflect variations of script shapes that help to visually distinguish a target script from others. Thus, Japanese kanji and Chinese
characters can be efficiently distinguished from other scripts. In the Japanese orthography, because of the average ratio of 30 percent kanji in sentences, kanji is generally easy to distinguish from hiragana and katakana, because of the larger stroke numbers.

Based on the numbers of lines (these are not the same as stroke numbers) used for constructing kanji, Kawai (1966) suggested that the complexity of a kanji character and its printed frequency approximately represent a log-linear relationship, described by \( C = 7.3 \log X + 0.1 \), where \( C \) is the complexity of the kanji and \( X \) is a ranking of the kanji's frequency of use (note that small \( X \) indicates a high frequency). This finding indicates that generally more frequent kanji tends to be simpler in structure. It should be noted that this log-linear relationship is a rough estimate for the overall understanding of kanji characters. However, this argument may support that complexity limits the sense-determinative function while enhancing the sense-discriminative function.

In fact, a more efficient sense-discriminative function does not imply a more efficient sense-determinative function. Imada and Yodogawa (1982, Experiment 1) found that the capital letters of the English alphabet took approximately 15 milliseconds per item, whereas the 1,945 *Jouyou Kanji* characters took 25–30 milliseconds per item. The difference of 10–15 milliseconds per item indicates that the English capital letters were perceived faster and more accurately than the 1,945 *Jouyou Kanji*. Furthermore, Imada and Yodogawa (1982, Experiment 2) measured the reaction times for simple kanji, intermediate kanji and complex kanji. The results indicated that the simple kanji were perceived in about 20 milliseconds, the intermediate kanji in 20–25 milliseconds, and the complex kanji in 30 milliseconds. These two experiments suggest that once the scripts become complicated
by their use of many lines, they become more visually multifarious. The simplicity of an alphabet, and probably kana, may contribute to producing faster and more accurate identification.

The sense-determinative function of kanji might depend on the 214 radicals or their combinations. Since a small number of radicals out of the 214 are repeatedly used in various kanji, the activation elements sufficient for accessing the kanji lexicon might be found in radicals or their combinations. Although kanji requires a higher average number of strokes compared to an alphabet and kana, many kanji alter their meanings depending on just a few strokes, such as 入 [/nyu/, enter] and 人 [/hito/, man] ; 水 [/mizu/, water] and 氷 [/koori/, ice]. This nature of kanji decreases the sense-discriminative function, but may increase the sense-determinative function (Kaiho & Nomura, 1983; Kaiho, 1987). Therefore, both functions must be blended in the scripts to an appropriate degree for efficient lexical access.

3.5 Summary

The six categories of kanji indicate that pictographic kanji comprise only about 10 percent of total kanji. On the other hand, ‘phonetic compound kanji’ constitute approximately on half of them. This figure implies that at least 50 percent of kanji could include both phonetic and visual cues. Furthermore, since kanji are constructed mostly by 214 radicals or their combinations, the radicals may act as critical activation elements to process kanji for lexical access. However, combinations of radicals and elements constructing complex kanji characters do not always indicate their meanings. The kanji content ratio in Japanese sentences is estimated at about 30 percent, which could be an optimal incidence to visually distin-
guish kanji from hiragana and katakana and aid in efficient reading.

The sense-determinative function and sense-discriminative function of the Japanese orthography suggest that, if only consider at a single symbolic level (not as a word unit), a single kana can be perceived faster than a single kanji, since kana symbols are simpler or more highly sense-determinative than kanji characters. Kanji, on the other hand, tend to contain high sense-discriminative functions due to the complex figures of script which can transfer richer meanings than kana. However, the more complex kanji becomes, the slower the recognition time. In other words, the complex figures of kanji can increase a degree of sense-discriminative function, but it is not necessarily associated with high sense-determinative function.

References


Imada, T., & Yodogawa, H. (1982). *Kanji no fukuzatsusa to rinkakusen keisei shori [Complexity of kanji and information processing in kanji frameworks]*. Showa 57
Nendo Denshi Tsushin Gattukai Sougo Zenkoku Taikai Happhyo Shoroku.

Tokyo: Taishuukan.

Gengo Seikatsu [Linguistic Life], 326, 68-79.

Tokyo: Koudansha.

In A. Sakakura (Ed.), Nihongo no rekishi [The history of Japanese] (pp. 113-154).
Tokyo: Taishuukan.


Tokyo: Akiyama Shoten.

Tokyo: Kyouiku Shuttupan.

Kawai, Y. (1966). Kanji no butsuriteki fukuzatsusei to yomi no gakushu [Physical complexity of Chinese characters and learning to read them].


Tokyo: Iwanami Shoten.

In H. Kindaichi (Ed.), Nihongo no sugata [Figure of Japanese] (pp. 4-37).
Tokyo: Taishuukan.

Tokyo: Koodansha.

Tokyo: Iwanami Shoten.

Tokyo: Meiji Tosho.

Tokyo: Shuei Shuppan.

Leong, C. K. (1973). Hong Kong. In J. Downing (Ed.), Comparative reading: Cross-


