

A Web-accessible database of characteristics of the 1,945 basic Japanese kanji

KATSUO TAMAOKA

Hiroshima University, Higashihiroshima, Japan

KIM KIRSNER

University of Western Australia, Perth, Australia

YUSHI YANASE

Ehime University, Matsuyama, Japan

YAYOI MIYAOKA

Hiroshima University, Higashihiroshima, Japan

and

MASAHIRO KAWAKAMI

Nagoya University, Nagoya, Japan

In 1981, the Japanese government published a list of the 1,945 basic Japanese kanji (Jooyoo Kanji-hyo), including specifications of pronunciation. This list was established as the standard for kanji usage in print. The database for 1,945 basic Japanese kanji provides 30 cells that explain in detail the various characteristics of kanji. Means, standard deviations, distributions, and information related to previous research concerning these kanji are provided in this paper. The database is saved as a Microsoft Excel 2000 file for Windows. This kanji database is accessible on the Web site of the Oxford Text Archive, Oxford University (<http://ota.ahds.ac.uk>). Using this database, researchers and educators will be able to conduct planned experiments and organize classroom instruction on the basis of the known characteristics of selected kanji.

Japanese kanji provides a stimulus-rich environment for research in the perceptual and cognitive processes required for reading, memory, and language acquisition in general. However, in order to do research on perceptual and cognitive processing using kanji, it is important to have a good understanding of the several potentially important differences between the Japanese writing system and other writing systems.

The first difference between the Japanese writing system and others is that three different scripts are used in Japanese: kanji, hiragana, and katakana. Kanji originally developed from pictographs used by the Chinese thousands of years ago to represent objects and events in the world around them. Some of these kanji have retained their pictographic forms and remain similar in appearance to the objects that they were intended to represent. Those kanji—such as 森, meaning *forest*, 門, meaning

gate, and 倉, meaning *storage*—that have maintained their pictographic image well in the process of script development may be easier for nonnative Japanese learners to acquire by using an imagery mediation strategy, as was indicated by Kuwabara (2000). Later on, other kanji were formed to represent more abstract ideas, whereas still others developed using combinations of *single* characters that convey information about a related idea. A fourth type of kanji consists of elements that are related to pronunciation. Hiragana and katakana are scripts representing morae (slightly smaller units within syllables) to depict the same set of 46 basic sounds. Hiragana is used for grammatical inflections and words where the writing of kanji is avoided. Katakana is used to write words and names that are not of Japanese or Chinese origin. No further consideration will be given to hiragana and katakana in this paper or in the associated database.

The second important difference between the Japanese writing system and others found in the world has to do with the number of basic kanji characters. In the version of kanji now in common use, there is a total of 1,945 basic characters. The difficulty associated with the mastery of this large set of characters is evident in the way the pedagogic load must be spread out over the first 9 years of schooling.

The authors thank Shoichi Yokoyama and Sanseido (publishing company) for allowing the use of the kanji frequency data from Yokoyama, Sasahara, Nozaki, and Long (1998). In addition, much gratitude is given to the reviewer, Yasushi Hino, as well as to an anonymous reviewer of this paper, for their valuable suggestions. Correspondence concerning this article should be addressed to K. Tamaoka, International Student Center, Hiroshima University, 1-1, 1-chome, Kagamiyama, Higashihiroshima, Japan 739-8523 (e-mail: ktamaoka@hiroshima-u.ac.jp).

A third difference associated with the Japanese writing system is the fact that kanji are basically constructed from a set of elements including 214 constituents, or *radicals*, which are used for classifying kanji in a kanji dictionary. With this being the case, all the elements making up each of the 1,945 basic kanji include one of these 214 radicals. There is a parallel in Indo-European writing systems, where many words have evolved from the same stem or root from a remote language, such as Latin. However, the radicals in kanji are not usually associated with particular sounds, as are those in alphabetic languages.

A fourth difference involves the fact that the pronunciation of an individual kanji character is often shared by several other kanji and, therefore, creates a large number of homophones. The English parallel, although with a much lesser occurrence, involves words such as BARE and BEAR, which, unlike most kanji examples, are visually similar.

A fifth difference involves the fact that many kanji (about 60% of the 1,945 basic Japanese kanji) have two or more pronunciations. These pronunciations are based on their two readings, one of Chinese origin (On-reading) and one of Japanese origin (Kun-reading).

The kanji database described and illustrated herein was created to facilitate general access to information about the many different characteristics of kanji and, in turn, to facilitate research in the perceptual and cognitive processes that may or may not be unique to kanji.

Kanji

In 1981, the Ministry of Education, Science, Sports, and Culture, Government of Japan (hereafter, simply called the *Japanese Ministry of Education*, except in quotes) published a list of the *1,945 Basic Japanese Kanji*, called “Jooyoo Kanji Hyoo” (常用漢字表; for detailed information, see Kato, 1989; Yasunaga, 1981).¹ This list established a standard for kanji usage in printed and written Japanese texts (Ministry of Education, Science, Sports, and Culture, Government of Japan, 1987, 1998). According to a survey conducted by the National Language Research Institute (1976) on the frequency of kanji in print, 2,000 kanji encompassed 99.6% of all the kanji used in three major Japanese newspapers (*Asahi*, *Mainichi*, and *Yomiuri*) published during the year of 1966. Although the 1,945 basic Japanese kanji and the 2,000 kanji mentioned above were not identical in each case, it is estimated that the basic kanji cover approximately 99% of all kanji used in Japanese newspapers.

In 1989, the Japanese Ministry of Education released a revised version of the Japanese language curriculum, “Nihongo Gakushuu Shidoo Yooryoo” (日本語学習指導要領), which included a list of kanji to be taught in Grades 1 to 6 (i.e., Gakunenbetsu Kanji Haitoohyo, 学年別漢字配当表). Of the 1,006 kanji on the list, 80 are to be taught in the first grade, 160 in the second grade, 200 in both the third and the fourth grades, 185 in the fifth

grade, and 181 in the sixth grade. All of these 1,006 kanji were taken from the 1,945 basic kanji. The remaining 939 kanji are taught in Grades 7 to 9. Because the Education Act in Japan stipulates that all Japanese citizens complete the ninth grade, it can therefore be assumed that native Japanese speakers educated in Japan study the 1,945 basic kanji and know them all by this grade.

Kanji characters, all of them having a meaning on their own, often combine to form various words with new meanings. For example, Yokosawa and Umeda (1988) found that approximately 70% of the 51,962 words listed in a particular Japanese dictionary were composed of two kanji combinations. Owing to the unique characteristics of this script, its 1,945 basic kanji are ideal for use in experiments and studies on cognitive processing. The present database provides 30 cells that cover various aspects of the fundamental characteristics of the 1,945 basic Japanese kanji. This information is stored as a Microsoft Excel 2000 file. By using this database, researchers will be able to conduct planned experiments based on the detailed information provided regarding the characteristics of each selected kanji.

The Kanji Table

This database includes 30 cells describing each of the 1,945 basic Japanese kanji. The cells, from left to right, indicate (1) the identification classification according to the Japanese 50-Sound System, “50-Onzu” (五十音図); (2) actual kanji orthography; (3) classification based on the six categories of “Rikusho Bunrui” (六書分類) provided by Shirakawa (1994); (4) school grade during which the kanji is taught; (5) which of the four levels of the Japanese Language Proficiency Test, “Nihongo Nooryoku Shiken” (日本語能力試験), each kanji is tested on; (6) the number of strokes to write the kanji; (7) kanji frequency provided by the National Language Research Institute (1976); (8) kanji frequency published by Yokoyama, Sasahara, Nozaki, and Long (1998); (9) kanji frequency on CD-ROM provided by Yokoyama et al.; (10) lexical productivity of the left-hand-side kanji provided by Kawakami (1997); (11) lexical productivity of the right-hand-side kanji provided by Kawakami; (12) the total lexical productivity of both the left-hand and the right-hand sides; (13) accumulative frequency of words produced by the left-hand-side kanji; (14) accumulative frequency of words produced by the right-hand-side kanji; (15) the total accumulative frequency of words produced by the left-hand and right-hand sides; (16) name of radicals used in the kanji; (17) radical frequency in the 1,945 basic kanji; (18) number of constituents that construct the kanji; (19) number of kanji homophones; (20) number of On-readings; (21) On-reading pronunciations; (22) inclusion of the three special Japanese sounds of /N/ (nasal), /Q/ (geminate), and /R/ (long vowel) in On-readings; (23) English translations of On-readings; (24) number of Kun-readings; (25) Kun-reading pronunciations; (26) inclusion of the three Japanese special sounds in Kun-

readings; (27) English translations of Kun-readings; (28) On-reading frequency calculated from the index provided by the National Language Research Institute (1976); (29) On- and Kun-reading frequency calculated from the index by the National Language Research Institute (1976); and (30) On-reading ratio calculated by dividing Cell 28 by Cell 29. Table 1 explains the stored data of all 30 cells in the kanji database.

Explanation of the 30 Cells

Cell 1. Kanji identification number as it ranks in the Japanese 50-Sound System.

Cell 2. Kanji character. A presentation of the character's orthography is provided here.

Cell 3. Kanji classification. The most common classification system for kanji is one developed by the Chinese scholar Kyoshin,² 許慎, who classified 9,353 kanji into the six categories of "Rikusho Bunrui" in the book entitled *Setsumon Kaiji* (説文解字; Atsui, 1988; Okai, 1934; Shirakawa, 2000a, 2000b; Tamaoka, 1991; Tanaka, 1982; Tomita & Sanada, 1994). This categorization system was later adopted by the Japanese. These six categories are (1) pictographic kanji (Shoeki, 象形), derived from the shapes of objects, (2) ideographic kanji (Shiji, 指事), which express ideas and qualities, (3) compound ideographic kanji (Kaii, 会意), formed by combining more than one internal constituent to represent ideas and their associations, (4) phonetic compound kanji

(Keisei, 形声), constructed by phonetic and semantic components, (5) loan kanji (Kashaku, 假借), whose original sounds were adopted, but not their original meanings, and (6) analogous kanji (Tenchuu, 轉注), which are new kanji patterned after old kanji to denote new meanings. To represent these categories, our Cell 3 used these six terms: 象形 (pictographic), 指事 (ideographic), 会意 (compound ideographic), 形声 (phonetic), 假借 (loan), and 轉注 (analogous). Analogous kanji are not found in the 1,945 basic Japanese kanji. These six categories were based on a system of categorization provided by Shirakawa (1994). It should be noted that there are five kanji³ (i.e., 込, 峠, 畑, 匆, and 粹) that cannot be classified according to this system of categorization. These kanji are original Japanese characters (Kokuji, 国字). They are labeled as 国字 (original). The kanji 堀 is an original Japanese character, but it is also a compound phonetic kanji, so it is listed as 形声 (phonetic).

As is shown in Table 2, unlike the common perception of kanji as being logographic or pictographic (e.g., Gelb, 1964; Hughes, 1962; Nishida, 1989), which classifies all kanji into only one of the six categories of Rikusho Bunrui, pictographic kanji were actually found in only 250 kanji out of the 1,945 basic kanji (12.85%). In contrast, it was found that compound phonetic kanji make up 59.85%, or 1,164, of the 1,945 basic kanji.⁴ Accordingly, Morioka (1988) called kanji *graphophonemes* or *morphophonemes*, which emphasizes the importance of the

Table 1
Explanation of Variables Stored in the 30 Cells of the Kanji Database

Cell #	Labels of Variables	Explanation of Variables	Example
1	ID	kanji identification number	48
2	Kanji	actual kanji orthography	右
3	Kanji Clas.	"Rikusho Bunrui" kanji classification	会意 Com. Ideographic
4	Grade	school grade in which the kanji is taught	1
5	JLPT-test	level of Japanese Language Proficiency Test	4
6	# of Strokes	number of kanji strokes	5
7	KF in 1976	kanji frequency provided by the National Japanese Research Institute (1976)	0.491
8	KF in 1998	kanji frequency provided by Yokoyama, Sasahara, Nozaki, and Long (1998)	7.407
9	KF on CD 1998	kanji frequency stored on CD-ROM provided by Yokoyama (1976)	8,971
10	Left Kanji Productivity	number of two-kanji compound words produced by a left-hand kanji	28
11	Right Kanji Productivity	number of two-kanji compound words produced by a right-hand kanji	4
12	Total Kanji Productivity	total number two-kanji compound words produced by both left and right kanji	32
13	Acc. Freq. Left Kanji Prod.	accumulative frequency of two-kanji compound words produced by a left kanji	57
14	Acc. Freq. Right Kanji Prod.	accumulative frequency of two-kanji compound words produced by a right kanji	44
15	Total Acc. Freq. Prod.	total frequency of two-kanji compound words produced by both left and right kanji	101
16	Name of Radical	name of radical	Kuchi
17	Radical Freq.	frequency of radicals found within the 1,945 basic Japanese kanji	38
18	# of Const.	number of kanji constituents	2
19	# of Homoph.	number of kanji homophones	4
20	# of On	number of On-readings	2
21	On-readings	On-readings	u, yuu
22	Sp. On	special sounds, /N/, /R/, and /Q/ in On-readings	R
23	Translation of On-readings	English translation of On-readings	right
24	# of Kun	number of Kun-readings	1
25	Kun-readings	Kun-readings	migi
26	Sp. Kun	special sounds, /N/, /R/, and /Q/ in Kun-readings	-
27	Translation of Kun-readings	English translation of Kun-readings	right
28	On-reading Freq.	kanji frequency of On-readings	157
29	On- & Kun-reading Freq.	total kanji frequency of both On- and Kun-readings	446
30	On-reading Ratio	On-reading ratio (Cell 28/Cell 29)	35.20%

Table 2
Numbers and Percentages of the 1,945 Basic Japanese Kanji
According to the “Rikusho Bunrui” Kanji Classification

Kanji Classification				
English	Japanese		Numbers	Percentage
Compound phonetic kanji	形声	Keisei	1,164	59.85
Compound ideographic kanji	会意	Kaii	503	25.86
Pictographic kanji	象形	Shookei	250	12.85
Loan kanji	仮借	Kashaku	12	0.62
Ideographic kanji	指事	Shiji	11	0.57
Japanese original kanji	国字	Kokuji	5	0.26
Analogous kanji	転注	Tenchuu	0	0.00
Total			1,945	100.00

Note—“Japanese original kanji” is not included in the six categories of the “Rikusho Bunrui.”

phonological aspect of kanji. In the experimental studies on the phonological processing of the Japanese language, Sakuma, Sasanuma, Tatsumi, and Masaki (1998) and Wydell, Patterson, and Humphreys (1993) provided evidence that both kanji orthography and phonology play an important role in the comprehension of kanji compound words. The second largest category was compound ideographic kanji, consisting of 503 kanji, or 25.86% of the basic kanji. Thus, 1,667 of the 1,945 basic kanji, or 85.71%, are classified as either compound phonetic kanji or compound ideographic kanji.

Cell 4. School grade of instruction. This cell specifies the school grade in which a kanji is taught. The assignment of the first 1,006 basic kanji to each school grade follows the Japanese language curriculum as outlined by the Japanese Ministry of Education in 1989. Since the remaining 939 basic kanji are taught in Grades 7–9, these are all indicated with the number 7. The kanji assigned to each grade is shown in Table 3. This index of assignment could be used as a means of finding the average age of acquisition of a particular kanji by native Japanese speakers or the school grade in which a kanji is taught. Pearson’s correlation coefficients ($n = 1,945$) between school grades and the indexes of kanji frequency were significantly high ($p < .01$), with the index of kanji fre-

quency in 1976 at $-.54$, in 1998 at $-.52$, and on CD-ROM in 1998 at $-.53$. Kanji acquired in the lower grades seem to be used more often in newspapers.

Cell 5. Kanji assignment to the four levels of the Japanese Language Proficiency Test. This test is prepared and administered by both the Japan Foundation (Kokusai Kooryuu Kikin) and the Association of International Education, Japan (Nihon Kokusai Kyooiku Kyookai) for nonnative Japanese speakers. In 1993, these two organizations provided test content specifications, one of which outlined kanji assignment to the four levels of difficulty (Aruku Nihongo Shuppan Henshuubu, 2000; Japan Foundation & Association of International Education, 1996). The kanji database indicates the levels of difficulty from the easiest of 4 to the most difficult of 1. It should be noted that 19 of the 1,945 basic Japanese kanji are not included in the Japanese Language Proficiency Test. Instead, the test includes an additional 110 kanji (a total of 2,036 kanji). The number of kanji assigned to each test level are 80 kanji for the fourth level, 165 kanji for the third level, 755 kanji for the second level, and 1,036 kanji for the first level. Thus, the kanji database can be used for selecting kanji for experiments or tests on nonnative Japanese speakers learning the Japanese language. Table 3 indicates the number of kanji

Table 3
The Number of Basic Japanese Kanji Required to be Taught in Each Grade
and for Each Level of the Japanese Language Proficiency Test
for Nonnative Japanese Speakers

School Grades	Total	Levels of Japanese Language Proficiency Test				Not Included
		4th Level	3rd Level	2nd Level	1st Level	
1	80	49	19	12	0	0
2	160	31	72	51	6	0
3	200	0	58	133	9	0
4	200	0	12	158	30	0
5	185	0	2	137	46	0
6	181	0	2	98	81	0
7–9	939	0	0	166	754	19
Total	1,945	80	165	755	926	19

Note—The levels of the Japanese Language Proficiency Test ranked from the easiest, 4, to the most difficult, 1.

in the proficiency test cross-matched with school grades. Pearson's correlation coefficient between school grades and levels of the proficiency test was significantly high ($n = 1,926, r = -.77, p < .01$),⁵ suggesting that kanji assignment to grades at Japanese schools is parallel to the assignment seen in the four levels of the Japanese proficiency test.

Cell 6. Number of strokes. This cell specifies the number of written strokes in each kanji. In 1958, the Japanese Ministry of Education published *Teaching Guidelines for Kanji Stroke Order*, which has been used continuously for over 40 years in Japanese schools (for detailed information, see Emori, 1988; Horiguchi, 1989; Matsumoto, 1998). These guidelines established two main rules and eight subrules for stroke order. The guidelines also provided proper stroke order for 881 kanji (the older version of the *Educational Kanji List*) to be learned from Grades 1 to 6. The number of strokes required to write a kanji were taken from a particular Japanese kanji dictionary edited by Kamata (1991), which followed these guidelines. Matsumoto explained that a few basic kanji have an alternative stroke order. These differences in stroke order occur because there are three coexisting sets of rules concerning stroke order: (1) the Japanese set of rules established in 1958, (2) the older set prior to 1958, and (3) the Chinese set of rules. It should be noted that Matsumoto suggested that these minor dif-

ferences should be considered equally viable and that kanji writing should not be restricted solely to the rules laid out by the Japanese Ministry of Education in 1958. Despite this suggestion, the present kanji database adopted the rules for stroke order established in 1958.

As is shown in Table 4, the mean number of strokes to write a kanji is 10.84, with a standard deviation of 3.76. The distribution of kanji strokes showed a relatively smooth normal curve with a skewness of 0.22 and a kurtosis of 2.76. Pearson's correlations ($n = 1,945$) between kanji strokes and the indexes of kanji frequency were moderate, with the index of kanji frequency in 1976 at $-.28 (p < .01)$, in 1998 at $-.27 (p < .01)$, and on CD-ROM in 1998 at $-.28 (p < .01)$. According to the correlation, the frequency of kanji is moderately related to the number of kanji strokes.

From their study on visual recognition of Chinese characters, Leong, Cheng, and Mulcahy (1987) stated that Chinese characters with many strokes took longer to visually recognize than did those with fewer strokes. In an experiment on the writing behavior of native Japanese speakers, Tamaoka and Takahashi (1999) measured the duration time for subjects to write two-kanji compound words from the moment the kanji were verbally presented to the subjects to the time kanji writing was initiated (i.e., writing latency). Their study indicated that high-frequency compound words with only a few strokes activated orthographic representations equally as fast as high-frequency kanji with many strokes. For low-frequency compound words, writing latency was longer for kanji with many strokes than for kanji with fewer strokes. The difference in writing latency may be caused by the stage of lexical access or by cognitive-motor planning to write the kanji. As such, number of strokes as an index of visual complexity seems to be an important factor for studies on lexical access and cognitive-motor planning.

Cell 7. Frequency of occurrence for kanji in print (kanji frequency) as indexed in 1976. Kanji frequency was calculated from all the words printed in three major newspapers (*Asahi*, *Yomiuri*, and *Mainichi*) during the year 1966 and was published by the National Language Research Institute in 1976. The study sampled 991,375 kanji and, for each one, calculated the frequency of appearance in every 1,000 kanji.

As is shown in Table 5, a majority of the 1,945 basic kanji range from 0.0 to 0.3 per million. The distribution of kanji frequency ($n = 1,925$, excluding the top 20 high-frequency kanji) is positively skewed, as is indicated, by 2.96 ($n = 1,945$). As is shown by a kurtosis of 10.25 ($n = 1,945$), a high peak appears in the lower range of kanji frequency, and it is gradually reduced to a high frequency range of over 3.0. The mean of the 1,945 basic kanji was 0.427, with a standard deviation of 0.459. However, the median of kanji frequency was far lower than the mean calculated as 0.156 per million ($n = 1,945$). The kanji with the highest frequency (i.e., mode) appeared 0.000 ($n = 1,945$) out of every 1,000 times.

Table 4
Distribution of the 1,945 Basic Japanese Kanji
(Based on Numbers of Kanji Strokes)

Stroke Numbers	Number of Kanji	Proportion (%)	Accumulative Proportion (%)
1	2	0.10	0.10
2	12	0.62	0.72
3	30	1.54	2.26
4	66	3.39	5.66
5	93	4.78	10.44
6	111	5.71	16.14
7	141	7.25	23.39
8	187	9.61	33.01
9	179	9.20	42.21
10	199	10.23	52.44
11	195	10.03	62.47
12	199	10.23	72.70
13	147	7.56	80.26
14	105	5.40	85.66
15	104	5.35	91.00
16	68	3.50	94.50
17	34	1.75	96.25
18	33	1.70	97.94
19	21	1.08	99.02
20	11	0.57	99.59
21	5	0.26	99.85
22	2	0.10	99.95
23	1	0.05	100.00
Total number of kanji		1,945	
Mean		10.84	
Standard deviation		3.76	
Skewness		0.22	
Kurtosis		2.76	
Median		10.76	
Mode		10	

Table 5
Distribution of 1,925 of the Basic Japanese Kanji
(Based on Kanji Frequency in 1976)

Range of Kanji Frequency	Number of Kanji	Proportion (%)	Accumulative Proportion (%)
0.0–0.3	1,232	64.00	64.00
0.3–0.6	280	14.55	78.55
0.6–0.9	150	7.79	86.34
0.9–1.2	77	4.00	90.34
1.2–1.5	49	2.55	92.88
1.5–1.8	34	1.77	94.65
1.8–2.1	26	1.35	96.00
2.1–2.4	17	0.88	96.88
2.4–2.7	16	0.83	97.71
2.7–3.0	10	0.52	98.23
3.0–3.3	8	0.42	98.65
3.3–3.6	7	0.36	99.01
3.6–3.9	7	0.36	99.38
3.9–4.1	6	0.31	99.69
4.1–4.4	3	0.16	99.84
4.4–4.7	3	0.16	100.00
Total number of kanji	1,925		
Mean	0.427		
Standard deviation	0.459		
Skewness	2.96		
Kurtosis	10.25		
Median	0.156		
Mode	0.000		

Note—The top 20 high-frequency kanji were excluded from the distribution. These kanji are listed in Index 1976 of Table 6.

The top 20 high-frequency kanji, which were excluded from the distribution of Table 5, are listed in Index 1976 of Table 6. The kanji with the highest frequency was 日 (/hi/), meaning *day*, which is naturally used for counting days and indicating the date of events. The kanji 一 (/iti/), meaning *one*, ranks second. These two kanji were the top two most frequently used kanji among the three indexes of kanji frequency stored in Cells 7–9.

The effects of kanji frequency were investigated by Tamaoka and Hatsuzuka (1995). They suggested that kanji frequency influences the processing of kanji compound words independently of word frequency. Similarly, frequency of Chinese characters was also found to affect Chinese word recognition (e.g., Taft, Huang, & Zhu, 1994; Taft & Zhu, 1995; Wu, Chou, & Liu, 1994; Zhang & Peng, 1992; Zhou & Marslen-Wilson, 1994). Thus, kanji frequency could be one of the major factors influencing cognitive processing of kanji compound words and should be controlled when conducting any experiments on the processing of a single or a compound word in kanji.

Cell 8. Kanji frequency as indexed in 1998. Yokoyama et al. (1998) published frequency of occurrence data based on all the kanji in the Tokyo edition of the *Asahi* newspaper printed in 1993. Their index recorded actual kanji frequency of occurrence 17,117,320 times (i.e., token frequency) for 4,583 kanji (i.e., type frequency). The mean frequency of occurrence for the 1,945 basic kanji was 7,316.07, with a standard deviation of 12,172.18. Although

the mean is relatively high, the median was 2,394, and the mode was very low at 98. Similar to the distribution of kanji frequency in 1976, kanji frequency in 1998 was also positively skewed ($n = 1,945$, a skewness of 2.92), with a high peak in the lower frequency range of 0–3,000 ($n = 1,945$, a kurtosis of 9.50), as is shown in Table 7. The distribution ranged from 0 to 3,000 ($n = 1,043$) and consisted of more than half of the kanji.

Pearson's correlation between frequency of occurrence for the 1,945 basic kanji ($n = 1,945$) in 1966, as recorded in our seventh cell, and in 1993, as recorded in our eighth cell, was found to be extremely high at .969 ($p < .01$). This figure indicates that the frequency of occurrence for printed kanji was stable over a period of 27 years, from 1966 to 1993. Likewise, Chikamatsu, Yokoyama, Nozaki, Long, and Fukuda (2000) found a significantly high correlation of .95 between the frequency of the 3,000 highest ranking kanji in the aforementioned index from 1993 and the index from 1966 (23 kanji in 1993 were not included in 1966). Therefore, the indexes of kanji frequency published in 1976 and in 1998 showed no difference concerning the 1,945 basic kanji, and even the 3,000 highest ranking kanji maintained similar frequency of occurrence over almost 3 decades.

Cell 9. Frequency of occurrence for kanji on CD-ROM. Yokoyama et al. (1998) calculated and published frequency of occurrence data for all the kanji used in the CD-ROM version of the *Asahi* newspaper (CD-HIASK'93). It contained 110,000 newspaper articles published in 1993. Thus, all the words in the Tokyo edition of the *Asahi* newspaper used in Cell 9 are included in the CD-ROM version. Again, this index used the actual frequency of kanji, indicated by the token frequency of 24,896,411 times for 4,488 kanji. The mean frequency among the 1,945 basic kanji was 12,484.16, with a standard deviation

Table 6
The Top 20 High-Frequency Kanji

1976 Index		1998 Index		1998 CD-ROM	
Kanji	Frequency	Kanji	Frequency	Kanji	Frequency
日	16,684	日	257,929	日	358,814
一	14,927	一	200,982	一	290,003
二	12,145	国	167,782	十	255,175
十	10,796	十	166,962	大	245,233
大	10,315	大	161,519	人	233,897
人	8,855	会	161,051	会	229,128
三	8,761	人	156,271	二	224,522
会	8,498	年	153,386	年	224,148
国	7,790	二	150,794	国	212,124
年	7,311	本	131,751	本	179,498
中	7,006	三	118,560	三	176,599
本	6,922	中	100,185	長	151,376
東	6,029	長	98,316	中	147,266
五	5,672	出	92,448	出	133,272
時	5,387	政	91,232	五	129,970
四	5,281	五	86,689	社	125,360
出	5,158	自	83,358	事	125,232
上	5,094	事	82,383	市	121,606
円	4,851	者	81,881	者	120,274
同	4,667	社	80,628	月	118,821

Table 7
Distribution of 1,925 of the Basic Japanese Kanji
(Based on Kanji Frequency in 1998)

Range of Kanji Frequency	Number of Kanji	Proportion (%)	Accumulative Proportion (%)
0-3,000	1,043	54.18	54.18
3,000-6,000	286	14.86	69.04
6,000-9,000	156	8.10	77.14
9,000-12,000	98	5.09	82.23
12,000-15,000	72	3.74	85.97
15,000-18,000	41	2.13	88.10
18,000-21,000	37	1.92	90.03
21,000-24,000	36	1.87	91.90
24,000-27,000	27	1.40	93.30
27,000-30,000	18	0.94	94.23
30,000-33,000	7	0.36	94.60
33,000-36,000	16	0.83	95.43
36,000-39,000	18	0.94	96.36
39,000-42,000	9	0.47	96.83
42,000-45,000	9	0.47	97.30
45,000-48,000	3	0.16	97.45
48,000-51,000	4	0.21	97.66
51,000-54,000	6	0.31	97.97
54,000-57,000	8	0.42	98.39
57,000-60,000	10	0.52	98.91
60,000-63,000	2	0.10	99.01
63,000-66,000	5	0.26	99.27
66,000-69,000	2	0.10	99.38
69,000-72,000	5	0.26	99.64
72,000-75,000	5	0.26	99.90
75,000-78,000	2	0.10	100.00
Total number of kanji	1,925		
Mean	7,316.07		
Standard deviation	12,172.18		
Skewness	2.92		
Kurtosis	9.50		
Median	2,394		
Mode	98		

Note—The top 20 high-frequency kanji were excluded from the distribution. These kanji are listed in Index 1998 of Table 6.

tion of 26,082.86. The median, however, was low at 3,631, and the mode was even lower at 119. Distribution was positively skewed ($n = 1,945$, a skewness of 5.51), with a very high peak ($n = 1,945$, a kurtosis of 44.34). Pearson's correlation for the relationship between kanji frequencies of occurrence established in 1966 and 1993 was very high at .971 ($p < .01$), suggesting that kanji frequency indexes changed little over 27 years. The correlation between the two kanji frequencies from the 1993 Tokyo editions of the *Asahi* newspaper and its CD-ROM version was .996 ($p < .01$). Therefore, although the CD-ROM version has more kanji than the actual newspaper in print, both the indexes of kanji frequency from Yokoyama et al. should reflect similar findings of kanji frequency for the 1,945 basic kanji.

Cell 10. Number of two-kanji compound words produced by a kanji in the left-hand side. Two-kanji compound word productivity refers to the unit of one kanji combined with another kanji to create two-kanji compound words. Therefore, the concept of this term could be understood as a linguistic concept of kanji *lexical productivity* (for

details, see Hayashi, 1987; Nomoto, 1989; M. Nomura, 1988, 1989). Thus, we call this the *lexical productivity of kanji*. This index was provided by Kawakami (1997). Two-kanji compound words are produced by the combination of kanji placed in the left-hand-side and right-hand-side positions of two-kanji compound words. For example, the kanji 学 (/gaku/, meaning *to learn*) is combined with another kanji in the right-hand-side position, such as in 学校 (/gakkō/, *school*), 学生 (/gakusei/, *student*), and 学者 (/gakusya/, *scholar*). Combinations with other kanji in the left-hand-side position are also possible, such as in 入学 (/nyūgaku/, *school admission*), 文学 (/bungaku/, *literature*), and 私学 (/sigaku/, *private school*). The number of lexical productivity is simply a count of all possible two-kanji combinations. The mean number of kanji lexical productivity in the left-hand side for each of the 1,945 basic kanji was 32.15, with a standard deviation of 40.86. The median was low at 20. The mode was 4. Distribution was positively skewed ($n = 1,945$, a skewness of 3.93), with a high peak ($n = 1,945$, a kurtosis of 26.11). Pearson's correlation between left (Cell 10) and right (Cell 11) kanji lexical productivity was .480 ($p < .01$). Psychological evidence of kanji activation as a morphemic unit is provided by Tamaoka and Hatsuzuka (1998), who suggested that the

Table 8
Distribution of the 1,945 Basic Japanese Kanji
(Based on Total Lexical Productivity of the
Left-Hand-Side and Right-Hand-Side Kanji)

Range of Word Production	Number of Kanji	Proportion (%)	Accumulative Proportion (%)
0-30	700	35.99	35.99
30-60	509	26.17	62.16
60-90	317	16.30	78.46
90-120	160	8.23	86.68
120-150	78	4.01	90.69
150-180	69	3.55	94.24
180-210	30	1.54	95.78
210-240	16	0.82	96.61
240-270	19	0.98	97.58
270-300	6	0.31	97.89
300-330	11	0.57	98.46
330-360	11	0.57	99.02
360-390	3	0.15	99.18
390-420	5	0.26	99.43
420-450	4	0.21	99.64
450-480	3	0.15	99.79
480-510	1	0.05	99.85
510-540	2	0.10	99.95
540-570	0	0.00	99.95
570-600	0	0.00	99.95
600-630	1	0.05	100.00
Total number of kanji	1,945		
Mean	64.64		
Standard deviation	70.34		
Skewness	2.73		
Kurtosis	10.28		
Median	42		
Mode	21		

Note—The *kanji lexical productivity* refers to how many two-kanji compound words are produced by a single kanji.

activation of semantic representations at the kanji level affects the processing of two-kanji compound words. Yet, it is still an open question as to how the size of kanji lexical productivity affects the cognitive processing of kanji compound words.

Cell 11. Lexical productivity of the right-hand-side kanji. The mean number of lexical productivity of the right-hand-side kanji for each of the 1,945 basic kanji was 32.25, with a standard deviation of 40.86. The median was low at 18. The mode was 1. Distribution was positively skewed ($n = 1,945$, a skewness of 2.97), with a high peak ($n = 1,945$, a kurtosis of 12.97). The numbers of two-kanji compound words produced by the right-hand-side kanji showed a similar distribution to the ones produced by the left-hand-side kanji, with a comparable mean and an identical standard deviation.

Cell 12. The total lexical productivity of both the left-hand- and the right-hand-side kanji. As is shown in Table 8, the mean total of lexical productivity in the left-hand- and right-hand-side kanji was 64.64, with a standard deviation of 70.34, for each of the 1,945 basic kanji. The median was slightly lower at 42, and the mode even lower at 21. As with the distributions of kanji lexical productivity in the right-hand and left-hand sides and the kanji frequency indexes, the distribution of total kanji lexical productivity was positively skewed ($n = 1,945$, a skewness of 2.73), with a high peak ($n = 1,945$, a kurtosis of 10.28). The high peak in the very positively skewed distribution suggests that 86.86% of kanji produce less than 120 two-kanji compound words. In fact, 1,209 kanji, or 62.16% of the 1,945 basic kanji, had fewer than 60 words. The top 20 kanji that produced the greatest number of two-kanji compound words are listed in Table 9. The kanji 大 (/dai/, big) had the greatest num-

ber of lexical productivity in the left-hand-side kanji, whereas the kanji 人 (/hito/ or /nin/, person) had the greatest number of lexical productivity in the right-hand side. As for total lexical productivity, the kanji 大 was the highest.

Cell 13. Accumulative lexical productivity of the left-hand-side kanji. Lexical productivity is simply a count of two-kanji compound words produced by a single kanji, with no consideration of word frequency. Grainger, O'Regan, Jacobs, and Segui (1989) found that visual recognition of a target word took longer when a neighboring word with a high word frequency was present (i.e., neighborhood frequency effect). They explained that lexical representations of neighboring words are activated simultaneously with those of the target word and compete with them. This competing process slows down the speed of visual recognition. In this sense, accumulative word frequency of all words together is considered to be more accurate in indicating the magnitude of the kanji lexical productivity effect than is a simple count of each produced word as 1. Thus, accumulative kanji lexical productivity is calculated by adding up all the frequency of occurrences for words in print provided by the National Language Research Institute (1973).

Since two-kanji compound words can be formed by the combining of kanji on the left-hand or the right-hand side of a targeted single kanji, Cell 13 provides the accumulative kanji lexical productivity of the left-hand side. The mean for accumulative lexical productivity of left-hand-side kanji was 57.26, with a standard deviation of 125.46, among the 1,945 basic kanji. The median was lower at 16, and the mode was at 0, since produced words had very low frequencies. The index in 1973 does not provide a word frequency lower than 5. In such a case,

Table 9
The Top 20 Kanji on the Basis of Lexical Productivity
(Number of Two-Kanji Compound Words and Their Accumulative Word Frequencies)

Lexical Productivity of Left-Side Kanji		Lexical Productivity of Right-Side Kanji				Total Lexical Productivity					
Kanji	Size	Kanji	Accum.	Kanji	Size	Kanji	Accum.	Kanji	Size	Kanji	Accum.
大	533	一	1,821	人	539	人	1,868	大	605	人	2,690
大小	434	大	1,766	子	429	十	1,427	大人	822	日	2,084
一	377	自	1,529	物	439	本	1,395	水	518	本	1,997
三	326	日	1,209	木	417	分	1,236	山	488	大	1,982
水	290	中	1,036	地	455	間	970	手	465	一	1,946
御	287	二	948	書	352	合	947	地	455	十	1,577
上	251	生	838	心	377	子	923	小	452	自	1,551
中	251	人	822	田	349	手	917	物	439	中	1,531
山	249	三	777	文	398	日	875	子	429	生	1,490
内	244	今	741	山	488	業	810	金	429	分	1,401
白	244	時	721	屋	264	事	802	一	427	事	1,399
手	233	小	665	道	321	田	777	木	417	手	1,361
本	230	場	664	手	465	年	771	中	412	場	1,351
下	229	国	663	水	518	場	687	本	408	子	1,250
新	225	主	623	行	335	学	677	文	398	国	1,216
金	222	最	616	事	267	度	669	上	394	合	1,171
高	221	以	605	色	306	画	667	心	377	学	1,121
天	217	本	602	金	429	画	662	下	375	会	1,119
神	210	発	599	目	302	生	652	石	366	会	1,061
地	205	事	597	頭	265	者	646	生	356	当	1,056

accumulative kanji lexical productivity becomes 0. There are many cases that fall into this type. In future kanji databases, possibly in the next edition, the accumulative kanji lexical productivity should be recalculated using a different word frequency index that provides more detailed word frequency information. The distribution for accumulative lexical productivity of the left-hand-side kanji was positively skewed ($n = 1,945$, a skewness of 6.19), with a high peak ($n = 1,945$, a kurtosis of 59.16). Pearson's correlation between accumulative kanji lexical productivity of left-hand-side (Cell 13) and right-hand-side (Cell 14) kanji was .377 ($p < .01$). Correlation between nonaccumulative (Cell 10) and accumulative left-hand-side kanji (Cell 13) was very high at .663 ($p < .01$).

Cell 14. Accumulative lexical productivity of the right-hand-side kanji. The mean for accumulative lexical productivity of right-hand-side kanji was 57.17, with a standard deviation of 129.14, among the 1,945 basic kanji. The median was lower at 14, and the mode was lower at 0 (see the explanation of Cell 13). As with the distribution for accumulative lexical productivity of left-hand-

side kanji, the distribution was also positively skewed ($n = 1,945$, a skewness of 5.59), with a high peak ($n = 1,945$, a kurtosis of 45.95). Pearson's correlation between the accumulative lexical productivity of left-hand- and right-hand-side kanji was .377 ($p < .01$). Correlation between nonaccumulative (Cell 11) and accumulative lexical productivity (Cell 14) was very high at .661 ($p < .01$).

Cell 15. Total accumulative lexical productivity of both the left-hand-side and the right-hand-side kanji. The mean for the total accumulative frequency of words produced by both left-hand- and right-hand-side kanji was 114.42, with a standard deviation of 214.94, among the 1,945 basic kanji. The median was lower at 42, and the mode was lower at 0 (see Cell 13). As is shown in Table 10, in which a sample of 1,925 kanji is used without the top 20 kanji shown in Index 1998 of Table 6, the distribution of total accumulative lexical productivity of both sides of kanji was also positively skewed ($n = 1,945$, a skewness of 4.71), with a high peak ($n = 1,945$, a kurtosis of 31.63). Pearson's correlation between nonaccumulative (Cell 12) and accumulative (Cell 15) total kanji lexical productivity was .713 ($p < .01$). Since the effects of kanji lexical productivity on the cognitive processing of kanji compound words is not clearly known as of yet, these figures may be useful for conducting studies on this topic.

Cell 16. Name of radical. This cell indicates the name of the radical used in each kanji. In 1615, a Chinese kanji dictionary called "Jii" (字彙) classified 33,197 Chinese characters on the basis of 214 basic elements. The majority of Japanese kanji dictionaries adopted this method and now classify Japanese kanji by way of these 214 radicals (for detail of these radicals, see Kaiho & Nomura, 1983; Kess & Miyamoto, 1999; Leong & Tamaoka, 1995; Saito, 1997; Saito, Masuda, & Kawakami, 1998; Shirakawa, 2000a, 2000b; Tamaoka, 1991; Todo, 1987, 1990). The name used for each radical has been taken from the Japanese kanji dictionary edited by Kamata (1991).

Cell 17. Radical frequency. Radical frequency indicates how many of the 1,945 basic kanji share the same radical. As is shown in Table 11, a great many kanji share the same small number of radicals out of the basic 214. In fact, only 24 radicals were used in constructing 54.34%, or 1,057, of the 1,945 basic kanji. Among these radicals, "Sanzui," representing *water*, was the most frequently found radical, used in more than 5.30% of the basic kanji, or 103 out of the 1,945 basic kanji. The radical "Ninben," which has the connotation of *man, human being, or person*, was the second most frequently used radical found in 87 kanji (4.47% of the basic kanji). The radical "Tehe" or "Te," which is connected to the meaning of *hand*, was the third most frequently used radical, found in 82 kanji. The top 10 most frequently used radicals are found in about 34.40%, or 669 out of the 1,945 basic kanji. Likewise, many characters in the Chinese language contain only a limited number of radicals. A study of 8,711 Chinese characters found that out of 214

Table 10
Distribution of 1,925 of the Basic Japanese Kanji
(Based on Accumulative Lexical Productivity of the
Left-Hand-Side and Right-Hand-Side Kanji)

Range of Accumulative Word Production	Number of Kanji	Proportion (%)	Accumulative Proportion (%)
0-50	1,051	54.60	54.60
50-100	327	16.99	71.58
100-150	177	9.19	80.78
150-200	97	5.04	85.82
200-250	74	3.84	89.66
250-300	42	2.18	91.84
300-350	27	1.40	93.25
350-400	20	1.04	94.29
400-450	25	1.30	95.58
450-500	13	0.68	96.26
500-550	8	0.42	96.68
550-600	8	0.42	97.09
600-650	16	0.83	97.92
650-700	8	0.42	98.34
700-750	3	0.16	98.49
750-800	8	0.42	98.91
800-850	8	0.42	99.32
850-900	3	0.16	99.48
900-950	4	0.21	99.69
950-1,000	4	0.21	99.90
1,000-1,050	1	0.05	99.95
1,050-1,100	1	0.05	100.00
Total number of kanji	1,925		
Mean	114.42		
Standard deviation	214.94		
Skewness	4.71		
Kurtosis	31.63		
Median	42		
Mode	0		

Note—The top 20 kanji with a high accumulative frequency of words produced by left-hand- and right-hand-side kanji are excluded from the distribution. These kanji are listed in the total kanji lexical productivity of Table 9.

Table 11
The 24 Most Frequently Used Radicals Among the 1,945 Basic Japanese Kanji

No.	Radicals in Kana	Radicals in Roma-ji	Meaning of Radicals	No. of Kanji	% of Kanji
1	サスイ	Sanzui	water	103	5.30
2	ニンベン	Ninben	man, human being	87	4.47
3	テヘン, テ	Tehen/Te	hand	82 (6)	4.22
4	キヘン, キ	Kihen/Ki	tree, wood	77 (23)	3.96
5	イトヘン, イト	Itohen/Ito	thread	60 (8)	3.08
6	ゴンベン	Gonben	speak, say	60	3.08
7	クチヘン, クチ	Kuchiheh/Kuchi	mouth	57 (38)	2.93
8	シンニョウ	Shin'nyou	advance, move	50	2.57
9	ヒヘン, ヒ	Hihen/Hi	fire	48 (27)	2.47
10	ツチヘン, ツチ	Tsuchiheh/Tsuchi	earth	45 (20)	2.31
11	ココロ	Kokoro	heart	40	2.06
12	クサカンムリ	Kusakanmuri	grass, plant	38	1.95
13	ウカンムリ	Ukanmuri	roof	36	1.85
14	オンナヘン, オンナ	On'nahen/On'na	woman	32 (8)	1.65
15	カネヘン, カネ	Kanehen/Kane	metal, gold	29 (1)	1.49
16	コサトヘン	Kozatohen	community	28	1.44
17	ニクヅキ	Nikuzuki	flesh, meat	28	1.44
18	リッシンヘン	Rissinben	heart	27	1.39
19	リットウ	Rittou	knife, sword	24	1.23
20	チカラ	Chikara	power	23	1.18
21	タケカンムリ, タケ	Takekanmuri/Take	bamboo	22 (1)	1.13
22	ノキヘン, ノキ	Nogihen/Nogi	grass, plant	22 (2)	1.13
23	コロモヘン, コロモ	Koromohen/Koromo	grain	20 (12)	1.03
24	マダレ	Madare	cliff	19	0.98
	Mean			43.31	2.26
	Standard deviation			23.68	1.22
	Total			1,057	54.34

Note—The numbers in parentheses indicate the second radical category. For example, the radical of Tehen and Te is found in 82 kanji, whereas the 6 in parentheses refers only to kanji with the radical Te. For the purpose of this study, both Tehen and Te are orthographically classified as a single radical, because they share the same origin.

radicals, 17 were utilized as basic elements in constructing 50.17% of these characters (Leong, 1973). Taking this linguistic evidence into account, it was assumed that knowledge relating to radicals would show a great effect on all aspects of the cognitive processing of Japanese kanji and Chinese characters.

Using the statistical method of path analysis, Tamaoka and Yamada (2000) found that knowledge of kanji radicals acts as a basic building block of lexical knowledge in all aspects of kanji orthography, phonology, and semantics. As previous studies have suggested (e.g., Leong & Tamaoka, 1995, Saito, 1997, and Saito et al., 1998, for Japanese kanji; Ho & Bryant, 1997, Ho, Wong, & Chan, 1999, Leong, 1973, and Taft & Zhu, 1995, 1997, for Chinese characters), radicals may serve as a fundamental element in the overall knowledge of kanji.

Cell 18. Number of perceptual kanji constituents. Single kanji are often composed of two or more constituents: a radical and secondary elements. We conducted a survey of visually complex kanji to identify how native Japanese speakers perceptually divide up a single kanji character. At the beginning of a general psychology class at Matsuyama University in Japan, a group of 226 university undergraduate students, all native speakers, were asked to divide up kanji, displayed in a large font on a piece of paper, into smaller constituents by circling them. There were 10 lists of kanji. Out of the

1,945 basic kanji, Lists 1–9 each contained 200 kanji, with List 10 having only 145 (a total of 1,945). Each subject completed only 1 of the 10 lists, and the number of subjects judging constituents in each kanji was controlled to no less than 20 and no more than 24. For data purposes, we rounded off the averages for the number of constituents to the nearest whole number. Our survey found that subjects were likely to divide up kanji first on the basis of radical units and then on the basis of other elements.

As is shown in Table 12, the mean number of kanji constituents was 2.19, with a standard deviation of 0.71, among the 1,945 basic kanji. The median and the mode were the same at 2. The distribution of kanji constituents

Table 12
Number of Kanji Constituents

Number of Kanji Constituents	Number of Kanji	Proportion (%)	Accumulative Proportion (%)
1	267	13.73	13.73
2	1,100	56.56	70.28
3	520	26.74	97.02
4	56	2.88	99.90
5	2	0.10	100.00
	Total	1,945	
	Mean	2.19	
	Standard deviation	0.71	

was a little positively skewed ($n = 1,945$, a skewness of 0.29), with a low peak ($n = 1,945$, a kurtosis of 0.15). Pearson's correlation between the number of kanji constituents (Cell 18) and the number of kanji strokes (Cell 6) was high at .641 ($p < .01$). The number of kanji constituents (Cell 18) was modestly correlated with the index of kanji frequency in 1976 (Cell 7) at $-.268$ ($p < .01$) and with the index of kanji frequency in 1998 (Cell 8) at $-.256$ ($p < .01$).

Cell 19. Number of kanji homophones. The Japanese borrowed their writing system of kanji mostly from the Chinese language but did not adopt the four tones that exist in standard Mandarin. This process increased the possibility of many kanji's sharing the same sound. In addition, kanji were borrowed from China during each period when Japan came into contact with China, and dialects varied depending on the capital city of the time period (for details of kanji adoption from China, see Hirai, 1979; Miller, 1967; Sato et al., 1978; Tamaoka, 1991). In other words, a single kanji that had already been adopted into the Japanese writing system may have had another pronunciation when it was introduced later on. This resulted in multiple pronunciations for a single kanji. The above-mentioned factor of ignoring the four Chinese tones increased the number of sounds shared by many kanji, and the factor of the different times of adoption increased the number of single kanji sharing different mul-

iple sounds. The combination of these factors provides the reason that a single pronunciation is often shared by many different kanji.

The number of kanji homophones was counted according to how many kanji out of the 1,945 basic kanji shared the same sound. For example, the sound /yoku/ can be written with five different characters of the 1,945 basic kanji. Each of these five kanji is indicated by the number 5. Both On- and Kun-readings were calculated for kanji homophones. When a single kanji has more than one pronunciation, the sound that has the greatest number of kanji homophones was recorded in Cell 19. The top 29 sounds shared by more than 20 kanji out of the 1,945 basic kanji are listed in Table 13. The most often shared sound was /syOR/, which was represented by 65 kanji (all On-reading) among the 1,945 basic kanji, or 3.34%. The sound /kor/, having 64 On-readings and 1 Kun-reading, also showed the same number of kanji homophony as the sound /syOR/. Almost all of these homophonic sounds are exclusively found in On-readings that were adopted from the Chinese sound system.

Among the 1,945 basic kanji, the mean number of kanji homophones was 17.37, with a standard deviation of 14.74. The median was at 12, and the mode was at 21. The distribution of kanji homophones is shown in Table 14. The figure showed a positively skewed distribution ($n = 1,945$, a skewness of 1.13), with a relatively high peak ($n =$

Table 13
The 29 Sounds Shared by More Than 20 Kanji
out of the 1,945 Basic Japanese Kanji

No.	Phoneme	No. of Kanji Homophones	Examples
1	/syOR/	65	小 称 生 症 勝 証 章 商 賞 省
2	/kor/	65 (one Kun-reading)	工 項 高 公 光 効 講 校 考 行
3	/si/	48	市 氏 死 資 子 士 史 師 視 志
4	/kan/	45	感 間 官 完 館 管 卷 觀 寒 関
5	/tor/	40	党 等 頭 当 刀 島 倒 投 統 討
6	/ka/	37 (three Kun-readings)	過 化 課 科 下 可 価 花 火 争
7	/sor/	35	層 荘 走 想 相 草 総 送 争 早
8	/ki/	35	氣 期 機 記 基 季 器 貴 危 企
9	/sei/	34 (one Kun-reading)	正 製 制 星 世 精 政 靜 成 整
10	/ken/	30	權 險 梟 軒 見 健 研 驗 圈
11	/kyOR/	30	強 教 橋 京 鏡 狂 經 競 協 共
12	/sin/	29	新 真 信 心 親 診 身 進 伸 深
13	/kei/	29	計 輕 系 型 景 係 經 敬 傾 繼
14	/tyOR/	26	超 町 丁 調 庁 兆 鳥 朝 張 頂
15	/sen/	26	先 干 戰 線 船 專 占 洗 選 宣
16	/sai/	25	再 最 際 才 歳 祭 探 裁 彩 細
17	/zyOR/	24	上 場 状 条 情 城 常 乘 冗 剃
18	/tei/	23	低 帝 定 訂 停 邸 提 庭 底 包
19	/hor/	23	方 法 報 訪 宝 胞 飽 放 袍 終
20	/sur/	23	週 州 集 周 宗 衆 秀 奧 普 浮
21	/fu/	23	不 會 回 界 解 階 海 介 開 改
22	/kai/	23 (one Kun-reading)	意 委 医 位 異 威 移 違 困 快
23	/i/	23 (one Kun-reading)	亡 坊 忘 望 翼 妨 暴 忙 防 房
24	/bor/	22	時 次 字 地 辟 持 兒 自 避 飛
25	/zi/	22 (one Kun-reading)	非 日 費 比 否 秘 被 陽 容 久
26	/hi/	21	用 要 幼 洋 祿 葉 被 陽 容 久
27	/yor/	21 (one Kun-reading)	旧 急 級 球 給 救 休 求 究 實
28	/kyUR/	20	对 隊 体 態 帶 退 耐 滯 待
29	/tai/	20	

Table 14
Distribution of the 1,945 Basic Japanese Kanji
(Based on Number of Kanji Homophony)

Range of Kanji Homophony	Number of Kanji	Proportion (%)	Accumulative Proportion (%)
0-3	183	9.41	9.41
3-6	267	13.73	23.14
6-9	278	14.29	37.43
9-12	193	9.92	47.35
12-15	165	8.48	55.84
15-18	66	3.39	59.23
18-21	100	5.14	64.37
21-24	170	8.74	73.11
24-27	102	5.24	78.35
27-30	57	2.93	81.29
30-33	62	3.19	84.47
33-36	67	3.44	87.92
36-39	37	1.90	89.82
39-42	0	0.00	89.82
42-45	44	2.26	92.08
45-48	45	2.31	94.40
48-51	49	2.52	96.92
51-54	0	0.00	96.92
54-57	0	0.00	96.92
57-60	0	0.00	96.92
60-63	60	3.08	100.00
Total number of kanji		1,945	
Mean		17.37	
Standard deviation		14.74	
Skewness		1.13	
Kurtosis		0.64	
Median		12	
Mode		21	

1,945, a kurtosis of 0.64). Pearson's correlation between the number of kanji homophones (Cell 19) and the number of kanji strokes (Cell 6) indicated no significant relation at $-.010$. The number of kanji homophones (Cell 18) had no correlation with the index of kanji frequency in 1976 (Cell 7) at $-.039$, the index of kanji frequency in 1998 (Cell 8) at $-.034$, the total kanji lexical productivity (Cell 12) at $-.048$, the total accumulative kanji lexical productivity (Cell 15) at $-.031$, and school grades (Cell 4) at $.006$. Even the phonological characteristic of On-

reading ratio (Cell 30) did not show any correlation with the number of kanji homophones ($r = .006$). Therefore, kanji homophony seems to be an independent characteristic.

Cell 20. Number of On-readings. Japanese kanji often have multiple On-readings. On-readings were adopted from the original Chinese pronunciation during the Chinese dynastic periods, when contact occurred between Japan and China. The count for On-readings was taken from the number of On-readings listed in the kanji dictionary edited by Kamata (1991), which followed the standards laid out by the Japanese Ministry of Education in 1981 for the 1,945 basic Japanese kanji. Number and percentage of kanji as cross-variables of the number of On- and Kun-readings among the 1,945 basic kanji are shown in Table 15. There were 777 out of 1,945 basic kanji (40.05%) that had only one type of either an On-reading or a Kun-reading: 40 kanji with only a Kun-reading and 737 kanji with only an On-reading. In other words, although it is a commonly held notion that a kanji has both an On-reading and a Kun-reading, only 1,166 kanji out of the 1,945 basic kanji, or 59.95%, had both types of pronunciations.

Cell 21. Pronunciation of On-readings. The pronunciation of On-readings is written in the Hepburn system of romanization provided by Nelson (1992). All the On-readings recorded in this cell are pronunciations based on the guidelines laid out by the Japanese Ministry of Education in 1981.

Cell 22. Special sounds /N/, /R/, and /Q/ in On-readings. The special sounds of /N/, /R/, and /Q/ are recorded in phonetic symbols of *N*, *R*, or *Q*, respectively, in this cell. The Japanese language during the Nara period (the years from 710 to 794) did not have any syllabic CVC combinations (Komatsu, 1981; Kubozono & Ota, 1998; Numoto, 1987). In modern Japanese, the two special sounds of /N/ for nasal and /Q/ for geminate frequently produce a syllabic structure of CVC, such as /ten/, /ken/, /haQ/, /kaQ/. These CVC-structured syllables began to appear in the Japanese sound system around the time of the

Table 15
On- and Kun-Readings for the 1,945 Basic Japanese Kanji

Number of Kun-Readings	Number of On-Readings						Total	Percentage
	0	1	2	3	4	5		
0	—	664	71	2	—	—	737	37.89
1	32	633	91	7	—	—	763	39.23
2	7	228	53	5	—	1	294	15.12
3	1	76	15	2	—	—	94	4.83
4	—	31	10	1	—	—	42	2.16
5	—	7	—	—	—	—	7	0.36
6	—	1	—	—	—	—	1	0.05
7	—	3	—	—	—	—	3	0.15
8	—	—	1	—	—	—	1	0.05
9	—	—	1	—	—	—	1	0.05
10	—	—	2	—	—	—	2	0.10
Total	40	1,643	244	17	0	1	1,945	100.00
Percentage	2.06	84.47	12.54	0.87	0	0.05	100.00	

Note—Rare pronunciations used for *Wago* and nouns of proper names and places were excluded.

Table 16
Special Sounds /N/, /R/, /Q/ Found in
On-Readings of the 1,945 Basic Japanese Kanji

Number of On-Readings	Type of Special Sounds	Number of Kanji
0	—	40
1	/R/	433
	/N/	359
2	None	851
	/R/ and /Q/	3
	/R/ and /N/	1
	/R/	91
	/N/	40
3	None	109
	/R/ and /Q/	2
	/R/ and /N/	1
	/R/	6
4	/N/	2
	None	6
5	—	0
5	/R/, /N/, and /Q/	1
Total		1,945
Kanji with special sounds		939
Kanji with no special sounds		966
Rate of occurrence of special sounds		49.29%

adoption of kanji from the Chinese language during and after the Heian period from 794 to 1192 (Kubozono & Ota, 1998). It should be noted that 40 kanji out of the 1,945 basic kanji do not have On-readings. Thus, only 1,905 kanji were taken into account. Out of the 1,905 basic kanji with On-readings, 939 contain the special sounds of /N/, /R/, or /Q/ in their On-readings, as is shown in Table 16. Since 966 out of the 1,905 kanji do not contain these special sounds, the rate of occurrence of special sounds found in On-readings is calculated as 49.29%.

Cell 23. English translation of On-readings.

Cell 24. Number of Kun-readings. Kun-readings originated in Japan. The count for the Kun-readings was taken from the kanji dictionary edited by Kamata (1991), which followed the guidelines of pronunciations specified by the Japanese Ministry of Education in 1981. Rare Kun-readings were not included in these guidelines.

Cell 25. Pronunciation of Kun-readings. Kun-readings were written in the Hepburn system of romanization provided by Nelson (1992). Again, all the Kun-readings recorded in this cell are pronunciations taken from the guidelines established by the Japanese Ministry of Education in 1981.

Cell 26. Special sounds /N/, /R/, and /Q/ in Kun-readings. As is indicated in Table 15, out of the 1,945 basic kanji, 737 kanji do not have Kun-readings. Thus, only 1,209 kanji were taken into account here. As compared with On-readings, only 48 with Kun-readings out of the 1,209 kanji contained a Japanese special sound (see Table 17). Thus, a majority of the kanji with Kun-readings (1,170 kanji) do not have special sounds in their pronunciations. The rate of occurrence of special sounds among kanji with Kun-readings is very low at 3.94%. This is a very interesting contrast to On-readings.

Cell 27. English translation of Kun-readings.

Cell 28. Accumulative frequency of occurrence of kanji with On-readings. Using the kanji frequency index provided by the National Language Research Institute in 1976, we calculated the frequency of occurrence for the On-readings of each kanji. The resulting figures give frequency of occurrence as it appeared in the three newspapers published in 1966. For example, the kanji 育, which has an On-reading of /iku/, appeared in 19 different words used 918 times in the sample taken from the newspapers. Accumulative frequency is therefore listed as 918. The names of people and places were not included in this frequency. When the overall accumulative kanji frequency was less than 9, the 1976 index did not provide detailed frequencies of either On- or Kun-readings. In this case, accumulative frequency is indicated as “—.”

Cell 29. Accumulative frequency of occurrence of both On- and Kun-readings. Following the kanji frequency index of 1976, a total accumulative frequency was calculated for each kanji, using both its On- and its Kun-readings. In the process of calculation, kanji used for proper nouns were excluded. This exclusion slightly alters the figures in this cell from the kanji frequency figures of Cell 7.

Cell 30. On-reading ratio. For our database, we calculated the On-reading ratio by dividing the accumulative On-reading frequency in Cell 28 by the total accumulative frequency for both On- and Kun-readings in Cell 29.

Table 17
Special Sounds /N/, /R/, and /Q/ Found in
Kun-Readings of the 1,945 Basic Japanese Kanji

Number of Kun-Readings	Type of Special Sounds	Number of Kanji
0	—	736
1	/R/	20
	/N/	5
	/Q/	5
2	None	743
	/R/	4
	/N/	2
3	None	288
	/R/	4
	/N/	1
	/Q/	1
4	/R/ and /N/	1
	None	87
	/Q/	1
	/N/ and /R/	3
	/N/ and /Q/	1
5	None	37
	None	7
6	None	1
7	None	3
8	None	1
9	None	1
10	None	2
Total		1,954
Kanji with special sounds		48
Kanji with no special sounds		1,170
Rate of occurrence of special sounds		3.94%

Note—Some kanji (9 kanji) with multiple pronunciations were counted twice. Thus, the total number of kanji with special sounds was 1,954.

Table 18
Distribution of 1,769 of the Basic Japanese Kanji
(Based on On-Reading Ratio)

Range of Kanji Homophony	Number of Kanji	Proportion (%)	Accumulative Proportion (%)
0-5	34	1.92	1.92
5-10	38	2.15	4.07
10-15	33	1.87	5.94
15-20	26	1.47	7.41
20-25	23	1.30	8.71
25-30	23	1.30	10.01
30-35	30	1.70	11.70
35-40	35	1.98	13.68
40-45	35	1.98	15.66
45-50	38	2.15	17.81
50-55	49	2.77	20.58
55-60	40	2.26	22.84
60-65	57	3.22	26.06
65-70	45	2.54	28.60
70-75	48	2.71	31.32
75-80	47	2.66	33.97
80-85	62	3.50	37.48
85-90	72	4.07	41.55
90-95	119	6.73	48.28
95-100	915	51.72	100.00
Total number of kanji		1,769	
Mean		79.12	
Standard deviation		8.43	
Skewness		-1.29	
Kurtosis		0.41	
Median		96.15%	
Mode		100%	

Note—Out of the 1,945 basic Japanese kanji, 176 kanji are excluded from the distribution above because 40 kanji have no On-readings (only Kun-readings) and 126 kanji have zero frequency in both On- and Kun-readings according to the kanji frequency index of 1976.

The On-reading ratios in the present database are objective figures using kanji frequency of occurrence. When a kanji has only one type of reading (either On or Kun), the figure is indicated as “-.” Among 1,769 of the basic kanji (176 kanji do not have more than one reading), the mean On-reading ratio was 79.12%, with a standard deviation of 8.43%. The median was very high at 96.15%, and the mode was very high at 100%. A distribution of kanji according to On-reading ratio is shown in Table 18. The figure showed a negatively skewed distribution ($n = 1,769$, a skewness of -1.29), with a relatively high peak ($n = 1,769$, a kurtosis of 0.41).

The correlation between the On-reading ratio (Cell 30) and the number of kanji homophones (Cell 19) was not significant at $-.006$, although both variables describe phonological aspects of kanji. The On-reading ratio had no correlation with the index of kanji frequency in 1976 (Cell 7) at $.031$ or with the one in 1998 (Cell 8) at $.045$. Thus, kanji frequency has nothing to do with On-reading ratio. The On-reading ratio had a very low correlation with total kanji lexical productivity (Cell 12) at $-.085$, total accumulative kanji lexical productivity (Cell 15) at $.062$, school grades (Cell 4) at $.110$, number of strokes (Cell 6) at $.112$, and number of constituents (Cell 18) at $.081$. All these correlations were at the 1% level of

significance, but the sample of 1,769 kanji was large enough that these correlations easily reach a significant level.

Some kanji have more than one On-reading. Using the Excel “sorting” command, a cross-variable table of kanji can be established on the basis of the number of On-readings (Cell 20) and the On-reading ratio (Cell 30). In this manner, various combinations of kanji variables can be examined by choosing targeted kanji.

According to the studies done by Y. Nomura (1978, 1979) and Kaiho and Nomura (1983), two variables are thought to distinguish On- and Kun-readings. The first variable is the degree of semantic concreteness. The studies done by Y. Nomura (1978, 1979) and Hirose (1998) suggest that kanji pronounced using Kun-readings are likely to be more independent and concrete in their meaning. The second variable is naming latency of kanji. In his experiments, Y. Nomura (1978, 1979) calculated On-reading ratio on the basis of kanji printed frequency provided by the National Japanese Research Institute (1962) and found that kanji with a low On-reading ratio (i.e., a high Kun-reading ratio) were pronounced with a longer latency than were kanji with a high On-reading ratio (i.e., a low Kun-reading ratio). He provided as an explanation that since kanji with Kun-readings are more concrete in their meaning, they cause an activation of semantic representations before the activation of phonological representations, even though the task required only vocalization of kanji pronunciation. Of course, as Y. Nomura (1978) added in his paper, there are some exceptional cases for this semantic attachment. For example, the kanji 菊, meaning *chrysanthemum*, is strongly associated with the sound /kiku/, which is an On-reading. Likewise, the kanji 肉 refers directly to the concrete concept of *meat*, but its pronunciation of /niku/ is an On-reading. It is reported that many native Japanese speakers confuse such exceptional kanji as being Kun-readings instead of On-readings (Kayamoto, 2000). Further research is needed to uncover how people distinguish On- and Kun-readings and even whether these distinctions are really important in the cognitive processing of kanji.

How to Obtain the Kanji Database

The database for 1,945 basic Japanese kanji is available on the Web site (<http://ota.ahds.ac.uk/>) of the Oxford Text Archive at Oxford University. The Oxford Text Archive was founded in 1976 by Lou Burnard in order to meet the research and teaching needs of electronic text users within the scholarly community. The Oxford Text Archive holds several thousand electronic texts and linguistic corpora, in a variety of languages, including electronic editions of works by individual authors, standard reference works such as the Bible, monolingual and bilingual dictionaries, and a range of language corpora.

The kanji database can also be acquired free of charge by contacting Katsuo Tamaoka through the regular mail (Katsuo Tamaoka, International Student Center, Hiroshima University, 1-1, 1-chome, Kagamiyama, Higashihiroshima,

Japan 739-8523) or e-mail (ktamaoka@hiroshima-u.ac.jp). The size of the database is 948KB stored as a Microsoft Excel 2000 file. Dr. Tamaoka will send the database stored on floppy disk by regular mail or attached to an e-mail message. Because no copyright protection is on this file, anyone can use the kanji database freely in the form of Excel 2000 or convert it to a text file, whatever the need may be. Besides giving credit to the authors of the database, there are no special requirements when using it for research purposes. It should be noted, however, that Japanese kanji cannot be displayed without Japanese Windows or a system with Japanese fonts.

REFERENCES

- ARUKU NIHONGO SHUPPAN HENSHUUBU [Department of Japanese Publishing, Aruku] (2000). *Nihongo nooryoku shiken kanji handobukku* [Kanji power handbook for the Japanese language proficiency test]. Tokyo: Aruku.
- ATSUJI, T. (1988). Kanji no bunrui: Rikusho wo chuushin toshite [Kanji classification: Rikusho six categories]. In K. Sato (Ed.), *Kanji kooza: Vol. 1. Kanji towa* [Kanji lecture series: Vol. 1. What is kanji?] (pp. 49-69). Tokyo: Meiji Shoin.
- CHIKAMATSU, N., YOKOYAMA, S., NOZAKI, H., LONG, E., & FUKUDA, S. (2000). A Japanese logographic character frequency list for cognitive science research. *Behavior Research Methods, Instruments, & Computers*, **32**, 482-500.
- EMORI, K. (1988). Jitai/shitsujun no shidoo [Teaching of script type and stroke order]. In K. Sato (Ed.), *Kanji kooza: Vol. 1. Kanji towa* [Kanji lecture series: Vol. 1. What is kanji?] (pp. 123-148). Tokyo: Meiji Shoin.
- GELB, I. J. (1964). *A study of writing*. London: University of Chicago Press.
- GRAINGER, J., O'REGAN, J. K., JACOBS, A. M., & SEGUI, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception & Psychophysics*, **45**, 189-195.
- HAYASHI, S. (1987). *Kanji, goi, bunshoo no kenkyuu e* [Studies of kanji, words and sentences]. Tokyo: Meiji Shoin.
- HIRAI, A. (1979). *Kokugo handobukku* [Japanese handbook]. Tokyo: Sanshodo.
- HIROSE, H. (1998). Identifying the On- and Kun-readings of Chinese characters: Identification of On versus Kun as a strategy-based judgment. In C. K. Leong and K. Tamaoka (Eds.), *Cognitive processing of the Chinese and the Japanese languages* (pp. 375-394). Dordrecht: Kluwer.
- HO, C. S.-H., & BRYANT, P. (1997). Learning to read Chinese beyond the logographic phase. *Reading Research Quarterly*, **32**, 276-289.
- HO, C. S.-H., WONG, W.-L., & CHAN, W.-S. (1999). The use of orthographic analogies in learning to read Chinese. *Journal of Child Psychology & Psychiatry*, **40**, 393-403.
- HORIGUCHI, J. (1989). Kanji no hitsujun [Stroke order of kanji]. In Y. Takebe (Ed.), *Nihongo to nihongo kyooiku: Dai-8-kan. Nihongo no moji hyooki* (Joo) [Japanese and Japanese education: Vol. 8. Japanese writing system, No. 1] (pp. 97-124). Tokyo: Meiji Shoin.
- HUGHES, J. P. (1962). *The science of language: An introduction to linguistics*. New York: Random House.
- JAPAN FOUNDATION AND ASSOCIATION OF INTERNATIONAL EDUCATION, JAPAN (1996). *Nihongo nooryoku shiken: Shutsudai kijun* [Japanese language proficiency test: Test content specifications]. Tokyo: Bonjinsha.
- KAIHO, H., & NOMURA, Y. (1983). *Kanji joochoo shori no shinrigaku* [Psychology of kanji information processing]. Tokyo: Kyoiku Shuppan.
- KAMATA, T. (1991). *Kuwashii shoogakkoo kanji jiten* [A detailed elementary school kanji dictionary]. Tokyo: Bun'eido.
- KATO, M. (1989). Kakushu kanji seigenan oyobi gen 'Jooyoo Kanji' wo meguru shojikoo ichiranhyoo [A proposal for the restriction of usage of various kanji and a list of information regarding the present 'Jooyoo Kanji']. In K. Sato (Ed.), *Kanji kooza: Vol. 11. Kanji to kokugo mondai* [Kanji lecture series: Vol. 11. Kanji and problems of the national language] (pp. 210-228). Tokyo: Meiji Shoin.
- KAWAKAMI, M. (1997). JIS 1-shu kanji 2965 ji wo mochiite sakusei sareru kanji niiji jukugosuu hyoo [Tables of two-kanji compound words constructed with 2,965 JIS 1st kanji characters]. *School of Education Bulletin* (Nagoya University), **44**, 243-299.
- KAYAMOTO, Y. (2000). Nihongo wo gakushuu suru chuugokugo bogo washa no kanji ninchi: Jookyuu/choojookyuu no naiteki jisho ni okeru on'in joochoo shori [Processing phonological information: Recognition of Japanese characters by advanced- and superior-level native speakers of Chinese]. *Kyoiku Shinrigaku Kenkyu* [Japanese Journal of Educational Psychology], **48**, 315-322.
- KESS, J. F., & MIYAMOTO, T. (1999). *The Japanese mental lexicon: Psycholinguistic studies of kana and kanji processing*. Amsterdam: John Benjamins.
- KOIZUMI, T. (1978). *Nihongo no seishohoo* [Japanese orthography]. Tokyo: Taishukan.
- KOMATSU, H. (1981). *Nihongo no on'in* [Japanese phonology]. Tokyo: Chuo Koronsha.
- KUBOZONO, H., & OTA, S. (1998). *On'in koozoo to akusento* [Phonological structure and accents]. Tokyo: Kenkyusha Shuppan.
- KUSAKABE, F. (1977). Nihon no romaji [The Roman alphabet in Japan]. In S. Ono & T. Shibata (Eds.), *Nihongo: dai 8 kan. Script* [Japanese: Vol. 8. Script] (pp. 340-383). Tokyo: Iwanami Shoten.
- KUWABARA, Y. (2000). Hikanjiken nihongo gakushuusha no kanji gakushuu ni okeru imeegi baikai hooryaku no yuukoosei [Imagery mediation strategy for learning Japanese characters: Paired-associate learning of characters and English translations]. *Kyoiku Shinrigaku Kenkyu* [Japanese Journal of Educational Psychology], **48**, 389-399.
- LEONG, C. K. (1973). Hong Kong. In J. Downing (Ed.), *Comparative reading: Cross-national studies of behavior and process in reading and writing* (pp. 383-402). New York: Macmillan.
- LEONG, C. K., CHENG, P.-W., & MULCAHY, R. (1987). Automatic processing of morphemic orthography. *Language & Speech*, **30**, 181-196.
- LEONG, C. K., & TAMAOKA, K. (1995). Use of phonological information in processing kanji and katakana by skilled and less skilled Japanese readers. *Reading & Writing*, **7**, 377-393.
- MATSUMOTO, H. (1998). Hitsujun-shi kenkyuu no koozoo [An idea of the study on the history of writing order]. *Bulletin of the Faculty of School Education, Hiroshima University* (Pt. II), **20**, 1-9.
- MILLER, R. A. (1967). *The Japanese language*. Chicago: University of Chicago Press.
- MINISTRY OF EDUCATION, SCIENCE, SPORTS, AND CULTURE, GOVERNMENT OF JAPAN (MONBUSHO) (1958). *Hitsujun shidoo no tebiki* [Teaching guidelines for kanji stroke order]. Tokyo: Hakubundo Shuppan.
- MINISTRY OF EDUCATION, SCIENCE, SPORTS, AND CULTURE, GOVERNMENT OF JAPAN (1987). *Shoogakkoo shidoosho: Kokugohen* [The Japanese language: The course of study at elementary school]. Osaka: Osaka Shoseki.
- MINISTRY OF EDUCATION, SCIENCE, SPORTS, AND CULTURE, GOVERNMENT OF JAPAN (1998). *Monbushoo kokuji: Shoogakkoo gakushuu shidoo yooryoo* [The announcement of the elementary school course of study by the Ministry of Education, Science, Sports, and Culture, Government of Japan.]. Tokyo: Gyosei.
- MORIOKA, K. (1988). Kanji no hyooisei to hyoo'onsei [Representing kanji semantics and phonology]. In K. Sato (Ed.), *Kanji kooza: Vol. 1. Kanji towa* [Kanji lecture series: Vol. 1. What is kanji?] (pp. 94-116). Tokyo: Meiji Shoin.
- NATIONAL JAPANESE RESEARCH INSTITUTE (1962). *Gendai zasshi 90 shu no yoogo yoorei* (Vol. 2) [Word usage in 90 current Japanese magazines]. Tokyo: Shuei Shuppan.
- NATIONAL LANGUAGE RESEARCH INSTITUTE (1973). *Shinbun no goi choosa* (IV) [A study of Japanese word usage in newspapers]. Tokyo: National Language Research Institute.
- NATIONAL LANGUAGE RESEARCH INSTITUTE (1976). *Gendai shinbun no kanji* [Japanese kanji characters in modern newspapers]. Tokyo: National Language Research Institute.
- NELSON, A. N. (1992). *The modern reader's Japanese-English character dictionary* (2nd rev. ed.). Tokyo: Charles E. Tuttle.
- NISHIDA, T. (1989). Sekai no moji [Writing systems in the world]. In

- T. Nishida (Ed.), *Sekai no moji* [Writing systems in the world] (pp. 5-41). Tokyo: Taishukan.
- NOMOTO, K. (1989). Miraishakai to kanji [Kanji in the future society]. In K. Sato (Ed.), *Kanji kooza: Vol. 11. Kanji to kokugo mondai* [Kanji lecture series: Vol. 11. Kanji and problems of national language] (pp. 210-228). Tokyo: Meiji Shoin.
- NOMURA, M. (1988). Niji kango no koozoo [Structure of two-kanji compound words]. *Nihongogaku* [Study on the Japanese language], **7**(5), 44-55.
- NOMURA, M. (1989). Kanji no zoogo ryoku [Productivity of kanji]. In K. Sato (Ed.), *Kanji kooza: Vol. 1. Kanji towa* [Kanji lecture series: Vol. 1. What is kanji?] (pp. 193-217). Tokyo: Meiji Shoin.
- NOMURA, Y. (1978). Kanji no joohoo shori: On-doku kun-doku no imi no fuka [The information processing of Chinese characters (kanji): Chinese reading, Japanese reading, and the attachment of meaning]. *Kyoiku Shinrigaku Kenkyu* [Japanese Journal of Psychology], **49**, 190-197.
- NOMURA, Y. (1979). Kanji no joohoo shori: On-doku kun-doku no kensaku katei [Information processing of Chinese characters (Kanji): Retrieval processes in Chinese-style reading (On) and Japanese-style reading (Kun)]. *Kyoiku Shinrigaku Kenkyu* [Japanese Journal of Psychology], **50**, 101-105.
- NUMOTO, K. (1987). *Nihon kanjion no kenkyuu* [A study on Japanese kanji sounds]. Tokyo: Tokyodo Shuppan.
- OKAI, S. (1934). *Nihon kanjigaku-shi* [Historical studies of Japanese kanji]. Tokyo: Meiji Shoin.
- SAITO, H. (1997). Shintekijisho [Mental lexicon]. In Y. Ohtsu & T. Gunji (Eds.), *Linguistic Sciences* (Vol. 3, pp. 93-153). Tokyo: Iwanami Shoten.
- SAITO, H., MASUDA, H., & KAWAKAMI, M. (1998). Form and sound similarity effects in kanji recognition. In C. K. Leong & K. Tamaoka (Eds.), *Cognitive processing of the Chinese and the Japanese languages* (pp. 169-203). London: Kluwer.
- SAKUMA, N., SASANUMA, S., TATSUMI, I. F., & MASAKI, S. (1998). Orthography and phonology in reading Japanese kanji words: Evidence from the semantic decision task with homophones. *Memory & Cognition*, **26**, 75-87.
- SATO, K., HACHIYA, K., KATO, M., HIDA, T., SATO, N., SUZUKI, T., & MAEDA, T. (1978). *Kokugogaku yosetsu* [A summarized explanation of the Japanese language]. Tokyo: Asakura Shoten.
- SHIRAKAWA, S. (1994). *Jitoo* [Kanji etymology]. Tokyo: Heibonsha.
- SHIRAKAWA, S. (2000a). *Kanji* [Kanji]. Tokyo: Iwanami Shoten.
- SHIRAKAWA, S. (2000b). *Shirakawa Shizuka chosaku: Shuu 3. Kanji III* [The series of articles by Shizuka Shirakawa: Vol. 3. Kanji III]. Tokyo: Heibonsha.
- TAFT, M., HUANG, J., & ZHU, X. P. (1994). The influence of character frequency on word recognition responses in Chinese. In H.-W. Chang, J.-T. Huang, C.-W. Hue, & O. J. L. Tzeng (Eds.), *Advances in the study of Chinese language processing* (Vol. 1, pp. 59-73). Taipei, Taiwan: Department of Psychology, National Taiwan University.
- TAFT, M., & ZHU, X. P. (1995). The representation of bound morphemes in the lexicon: A Chinese study. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 293-316). Hillsdale, NJ: Erlbaum.
- TAFT, M., & ZHU, X. P. (1997). Submorphemic processing in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 761-775.
- TAMAOKA, K. (1991). Psycholinguistic nature of the Japanese orthography. *Studies in Language and Literature* (Matsuyama University), **11**, 49-82.
- TAMAOKA, K., & HATSUZUKA, M. (1995). Kanji niji jukugo no shori ni okeru kanji shiyoo hindo no eikyoo [The effects of Kanji printed-frequency on processing Japanese two-morpheme compound words]. *Science of Reading*, **39**, 121-137.
- TAMAOKA, K., & HATSUZUKA, M. (1998). The effects of morphological semantics on the processing of Japanese two-kanji compound words. In C. K. Leong & K. Tamaoka (Eds.), *Cognitive processing of the Chinese and the Japanese languages* (pp. 139-168). London: Kluwer.
- TAMAOKA, K., & TAKAHASHI, N. (1999). Kanji niji jukugo no shoji koodoo ni okeru goi shiyoo hindo oyobi shojiteki fukuzatusei no eikyoo [The effects of word frequency and orthographic complexity on the writing process of Japanese two-morpheme compound words]. *Japanese Journal of Psychology*, **70**, 45-50.
- TAMAOKA, K., & YAMADA, H. (2000). The effects of stroke order and radicals on the knowledge of Japanese kanji orthography, phonology and semantics. *Psychologia*, **43**, 199-210.
- TANAKA, I. (1982). *Nihon no kanji* [Japanese kanji]. Tokyo: Chuo Koronsha.
- TODO, A. (1987). *Kanji no hanashi: Joo* [Story of kanji: Vol. 1]. Tokyo: Asahi Shinbunsha.
- TODO, A. (1990). *Kanji no hanashi: Ge* [Story of kanji: Vol. 2]. Tokyo: Asahi Shinbunsha.
- TOMITA, T., & SANADA, K. (1994). *Shin-Hyooki* [Revised: Orthography]. Tokyo: Kokusai Koryu Kikin.
- WU, J.-T., CHOU, T.-L., & LIU, I.-M. (1994). The locus of the character/word frequency effect. In H.-W. Chang, J.-T. Huang, C.-W. Hue, & O. J. L. Tzeng (Eds.), *Advances in the study of Chinese language processing* (Vol. 1, pp. 31-58). Taipei: Department of Psychology, National Taiwan University.
- WYDELL, T. N., PATTERSON, K. E., & HUMPHREYS, G. W. (1993). Phonologically mediated access to meaning for kanji: Is a rose still a rose in Japanese kanji? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 491-514.
- YASUNAGA, M. (1981). Jooyoo kanjihyoo ga umarerumade [A background history of the Jooyoo Kanji List]. *Gengo Seikatsu*, **355**, 24-31.
- YOKOSAWA, K., & UMEDA, M. (1988). Processes in human Kanji-word recognition. *Proceedings of the 1988 IEEE international conference on systems, man, and cybernetics* (pp. 377-380). Beijing and Chongyang.
- YOKOYAMA, S., SASAHARA, H., NOZAKI, H., & LONG, E. (1998). *Shinbun denshi media no kanji: Asahi shinbun CD-ROM niyoru kanji hindo hyoo* [Japanese kanji in the newspaper media: Kanji frequency index from the Asahi newspaper on CD-ROM]. Tokyo: Sanseido.
- ZHANG, B., & PENG, D. (1992). Decomposed storage in the Chinese lexicon. In H.-C. Chen & O. J. L. Tzeng (Eds.), *Language processing in Chinese* (pp. 131-149). Amsterdam: North-Holland.
- ZHOU, X., & MARSLER-WILSON, W. (1994). Words, morphemes and syllables in the Chinese mental lexicon. *Language & Cognitive Processes*, **9**, 393-422.

NOTES

1. The paper, in general, used the Hepburn style of romanization (i.e., *Roma-ji*) to describe Japanese words (for detail of the *Roma-ji* script, see Koizumi, 1978; Kusakabe, 1977). However, the Hepburn style does not distinguish short and long vowels. Thus, the present paper repeated vowels twice as /aa/, /ii/, /uu/, /ee/, and /oo/ in order to differentiate short vowels of /a/, /i/, /u/, /e/, and /o/. Names of authors quoted and referred to in this paper are written in the Hepburn style.

2. In this paper, the names of Chinese people and books are written in *Roma-ji* on the basis of pronunciations in Japanese.

3. The five Japanese original kanji (Kokuji) can be considered as compound ideographic kanji (Kaiji) because they were created from the meanings of already-existing kanji. For example, the kanji 畑 (/hatake/, a farming field) is produced by combining the two kanji 田 (meaning a field) and 火 (meaning fire) to represent how a traditional farming field is prepared by burning its weeds by fire. Likewise, the kanji 峠 (/torge/, a mountain pass) was formed by combining three kanji 山 (mountain), 上 (up), and 下 (down). This kanji, then, refers to the highest point of a mountain road.

4. Tomita and Sanada (1994) counted compound phonetic kanji "Keisei" as making up 1,286 out of the 1,945 basic kanji, or 66.12%. This figure is larger than the one in our kanji database, which adopted kanji categorization from Shirakawa (1994).

5. The sample for correlation between school grades and levels of the Japanese proficiency test used only 1,926 of the 1,945 basic kanji, since 19 kanji were not included in the outline of the kanji list for the Japanese proficiency test.