Behavior Research Methods, Instruments, & Computers
2004, 36 (3), 548-558

New figures for a Web-accessible database of the 1,945 basic Japanese kanji, fourth edition

KATSUO TAMAOKA
Hiroshima University, Hiroshima, Japan

and

SHOGO MAKIOKA
Osaka Women's University, Osaka, Japan

On the basis of calculations using the latest lexical database produced by Amano and Kondo (2000), the fourth edition of a Web-accessible database of characteristics of the 1,945 basic Japanese kanji was produced by including the mathematical concepts of *entropy*, *redundancy*, and *symmetry* and by replacing selected indexes found in previous editions (Tamaoka, Kirsner, Yanase, Miyaoka, & Kawakami, 2002). The kanji database in the fourth edition introduces seven new figures for kanji characteristics: (1) printed frequency, (2) lexical productivity, (3) accumulative lexical productivity, (4) symmetry for lexical productivity, (5) entropy, (6) redundancy, and (7) numbers of meanings for On-readings and Kun-readings. The file of the fourth edition of the kanji database may be downloaded from the Psychonomic Society Web archive, http://www.psychonomics.org/archive/.

In 1981, Japan's Ministry of Education, Culture, Science, Sports, and Technology (1987, 1998; hereafter, the *Japanese Ministry of Education*) established a standard for the usage of the 1,945 basic Japanese kanji in printed texts. This list is known as *Jooyoo Kanji Hyoo* (the list of commonly used kanji; for detailed information, see Kato, 1989; Yasunaga, 1981). Two decades later, Tamaoka, Kirsner, Yanase, Miyaoka, and Kawakami (2002) produced a Web-accessible database of characteristics of the 1,945 basic Japanese kanji. In 2004, the fourth edition of this evolving effort introduced new figures to replace or complement existing ones on the basis of calculations using the latest and most comprehensive lexical database produced by Amano and Kondo (2000). In addition to the use of the newer database, an important difference between the fourth and the previous editions is the inclusion of the mathematical indexes *entropy*, *redundancy*, and *symmetry*.

The concept of *neighborhood* in alphabetic languages has been extensively investigated (e.g., Andrews, 1989, Coltheart, Davelaar, Jonasson, & Besner, 1977, and Snodgrass & Mintzer, 1993, for English; Grainger, 1990, for Dutch; Grainger & Segui, 1990, and Grainger, O'Regan, Jacobs, & Segui, 1989, 1992, for French; Van Heuven, Dijkstra, & Grainger, 1998, for Dutch and English bilinguals). Orthographic neighbors of alphabetic words are defined as those that can be generated by replacing one letter with another while preserving the positions of the remaining letters. Since Japanese kanji do not have such alphabetic letter units, kanji neighbors can be understood as morphemic units that create various words. Since kanji morphemic units are too large to be defined as a single lexical neighborhood, the present database indicates the frequency counts of two-kanji compound words produced on the basis of the same single kanji on one side with the term *kanji lexical productivity*.

Mathematical indexes can be applied to detailed studies of kanji neighborhoods. The concept of *entropy* was first developed by the American mathematician Claude Elwood Shannon (1916–2001) in his seminal work, *A Mathematical Theory of Communication* (1948). Shannon explored entropy as an index for the degree of disorder or chaos (for elaboration, see Hori, 1979; Kaiho, 1989; Tamaoka, Miyaoka, & Lim, 2003). Applying this concept to kanji, entropy can provide the degree of irregularity (or regularity) of kanji's creating various two-kanji compound words in the left-hand and right-hand sides. Using kanji lexical productivity, a single value of entropy provides information as to the various overall distributions of neighboring lexical items. Since frequencies of neighborhood words alter the speed of word processing (Grainger, 1992), the index of entropy provides unique possibilities for exploring kanji neighborhood patterns. In addition, another of Shannon's mathematical concepts, *redundancy*, indicates repetition of lexical items in an overall distribution of lexical items produced by the same single kanji. The fourth edition of the kanji database therefore provides mathematically produced kanji indexes that make further detailed experiments possible for various studies in psychology, linguistics, psycholinguistics, and other related areas.

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).

Table 1
Explanation of Variables Stored in the 38 Rows of the Kanji Database, Fourth Edition

Row	Label of Variables	Explanation of Variables
1	ID	kanji identification number
2	Kanji	actual kanji orthography
3	Kanji Clas.	Rikusho Bunrui kanji classification based on Shirakawa (1994, 2003)
4	Grade	school grades in which each kanji is taught in Japanese schools
5	JLPT-test	level of Japanese Language Proficiency Test
6	# of Strokes	number of kanji strokes
7	KF in 1976	kanji frequency provided by the National Institute for Japanese Language (1976)
8	KF in 1998	kanji frequency provided by Yokoyama, Sasahara, Nozaki, and Long (1998)
9	KF on CD 1998	kanji frequency stored on CD-ROM provided by Yokoyama, Sasahara, Nozaki, and Long (1998)
10	KF in 2000	kanji frequency calculated using Amano and Kondo (2000)
11	Left Kanji Prod. 2000	kanji lexical productivity on the left-hand side calculated using Amano and Kondo (2000)
12	Right Kanji Prod. 2000	kanji lexical productivity on the right-hand side calculated using Amano and Kondo (2000)
13	Total Prod. 2000	total number of kanji lexical productivity on both left and right kanji calculated using Amano and Kondo (2000)
14	Acc. Freq. Left Prod. 2000	accumulative frequency of kanji lexical productivity on the left-hand side calculated using Amano and Kondo (2000)
15	Acc. Freq. Right Prod. 2000	accumulative frequency of kanji lexical productivity on the right-hand side calculated using Amano and Kondo (2000)
16	Total Acc. Freq. Prod. 2000	total frequency of kanji lexical productivity on the left- and right-hand sides calculated using Amano and Kondo (2000)
17	Symmetry 2000	symmetry of kanji lexical productivity on the left- and right-hand sides calculated using Amano and Kondo (2000)
18	Left Entropy 2000	entropy of kanji lexical productivity on the left-hand side calculated using Amano and Kondo (2000)
19	Right Entropy 2000	entropy of kanji lexical productivity on the right-hand side calculated using Amano and Kondo (2000)
20	Left Redundancy 2000	redundancy of kanji lexical productivity on the left-hand side calculated using Amano and Kondo (2000)
21	Right Redundancy 2000	redundancy of kanji lexical productivity on the right-hand side calculated using Amano and Kondo (2000)
22	Name of Radical	name of radical
23	Radical Freq.	frequency of radicals found within the 1,945 basic Japanese kanji
24	# of Const.	number of kanji constituents
25	# of Homoph.	number of kanji homophones
26	# of Meanings of On	number of meanings pronounced in On-readings
27	Translation of On-readings	English translation of On-readings
28	# of On	number of On-readings
29	On-readings	On-readings
30	Sp. On	special sounds, /N/, /R/ and /Q/ in On-readings
31	# of Meanings of Kun	number of meanings pronounced in Kun-readings
32	Translation of Kun-readings	English translation of Kun-readings
33	# of Kun	number of Kun-readings
34	Kun-readings	Kun-readings
35	Sp. Kun	Special sounds, /N/, /R/, and /Q/ in Kun-readings
36	On-reading Freq.	kanji frequency of On-readings calculated using the National Institute for Japanese Language (1976)
37	On- & Kun-reading Freq.	total kanji frequency of both On- and Kun-readings calculated using the National Institute for Japanese Language (1976)
38	On-reading Ratio	On-reading ratio (Cell 36/Cell 37) calculated using the National Institute for Japanese Language (1976)

Overview of the Kanji Database, Fourth Edition

As is shown in Table 1, rows of the kanji database in the fourth edition reflect a rearrangement for practical purposes of those in previous editions. Furthermore, the latest edition presents seven new figures for kanji characteristics: (1) kanji printed frequency in Row 10, (2) kanji lexical productivity in Rows 11–13, (3) accumulative kanji lexical productivity in Rows 14–16, (4) symmetry in Row 17, (5) kanji entropy in Rows 18 and 19, (6) kanji redundancy in Rows 20 and 21, and (7) numbers of meanings for On-readings in Row 26 and Kun-readings in Row 31. Since other figures and their related research issues were discussed at length in Tamaoka et al. (2002), the present study focused on the aforementioned new figures and certain calculation formulas.

Kanji Printed Frequency in Row 10

Amano and Kondo (2000) produced a large lexical database of printed frequencies for 341,771 words (i.e., type frequency) taken from editions of the *Asahi* newspaper printed from 1985 to 1998, containing 287,792,797 words (i.e., token frequency). Using this lexical database, kanji printed frequencies for 1,945 basic Japanese kanji were calculated by the programming language of MacJPerl 5.15r4J for Macintosh. The fourth edition kanji database includes this new index of kanji-printed frequencies. As is shown in Table 2, a large number of 1,417 kanji (72.85% of 1,945 kanji) had printed frequencies of occurrence ranging from 0 to 3,999 times. The new kanji frequency index recorded actual kanji printed frequency of occurrence for the 1,945 basic kanji to be a total of 86,542,349 times.

Table 2
Distribution of Kanji Printed Frequency for the 1,945 Basic Japanese Kanji

Range of Kanji Frequency	Number of Kanji	Percentage	Accumulative Percentage
0–39,999	1,417	72.85	72.8
40,000–79,999	210	10.80	83.65
80,000–119,999	118	6.07	89.72
120,000–159,999	59	3.03	92.75
160,000–199,999	42	2.16	94.91
200,000–239,999	21	1.08	95.99
240,000–279,999	19	0.98	96.97
280,000–319,999	15	0.77	97.74
320,000–359,999	11	0.57	98.30
360,000–399,999	10	0.51	98.82

Note—Figures beyond an accumulative percentage of 99% were excluded.

The distribution ranged from 0 to 920,283 times (the kanji 会 showing the highest frequency). The median of kanji distribution was 10,749 times. Since 11 kanji did not appear at all in the *Asahi* newspaper for 14 years, the most frequent occurrence of kanji printed frequency (i.e., mode) was 0 times. The mean was 44,495 times, with a standard deviation of 85,666 times. As is shown in Table 2, the kanji frequency distribution was positively skewed (skewness = 3.93) with a high peak (kurtosis = 21.00).

The second and third editions of the kanji database (for details, see Tamaoka et al., 2002) contained an index of kanji printed frequency from the three major daily newspapers (*Asahi*, *Yomiuri*, and *Mainichi*) provided by the National Institute for Japanese Language (1976), and two more indexes of kanji frequencies in both the printed and the CD-ROM versions of the Tokyo edition of the *Asahi* newspaper in 1993, as provided by Yokoyama, Sasahara, Nozaki, and Long (1998). Pearson's correlations of printed frequencies for the 1,945 basic kanji (n = 1,945) in the National Institute for Japanese Language (1976) were extremely high at .969 ($p < .01$) for the printed version and .971 ($p < .01$) for the CD-ROM version. The new kanji frequency index shows a slightly lower but, nevertheless, considerable correlation of .721 ($p < .01$) with the index of 1976: .799 ($p < .01$) for the printed version, and .789 ($p < .01$) for the CD-ROM version.

The 1,945 basic kanji appeared a total of 16,707,546 times in print and 24,281,697 times on CD-ROM in the 1998 indexes; however, the new kanji printed frequency index was calculated from a total of 86,542,349 times for the same number of kanji. Even though correlations were not as high as those between the indexes of 1976 and 1998, the size of the latest sample—approximately four times larger than that of the 1998 index—yields a better index for printed frequencies of the 1,945 basic kanji.

Kanji Lexical Productivity on the Left- and Right-Hand Sides in Rows 11–13

A single kanji can produce two-kanji compound words in two ways, through the combination of kanji placed on the left- and right-hand sides of two-kanji compound words. For example, the kanji 学 /gaku/ meaning "to learn" or "learning" is combined with another kanji on the right-hand side, such as in 入学 (/njur gaku/, "school admission"), 文学 (/bun gaku/, "literature"), and 私学 (/si gaku/, "private school"). Combinations with other kanji on the right-hand side are also possible, such as in 学校 (/gaq kor/, "school"), 学生 (/gaku sei/, "student"), and 学者 (/gaku sja/, "scholar"). Kanji productivity of two-kanji compound words refers to two units of kanji being combined to create two-kanji compound words. Therefore, this term could be understood as a linguistic concept of *kanji lexical productivity* (for details, see Hayashi, 1987; Nomoto, 1989; Nomura, 1988, 1989), although previous editions of the kanji database called this *kanji neighborhood size*. The fourth edition provides a number of two-kanji compound words (i.e., type frequency of kanji lexical productivity), counted using the lexical corpus of Amano and Kondo (2000). These figures are reported in Row 11 for the left-hand side, in Row 12 for the right-hand side, and in Row 13 for both sides together.

As is shown in Table 3, about half of the 1,945 kanji (931, or 47.87%) produced 0–9 two-kanji compound words on the left-hand side. Furthermore, about 90% of the 1,945 basic kanji showed fewer than 50 words having kanji lexical productivity. The kanji 人 (/zin/ in On-reading and /hito/ in Kun-reading, "human") showed the highest kanji lexical productivity on the left-hand side, producing 268 words. The second highest was 子 (/si/ in On-reading and /ko/ in Kun-reading, "child"), having 246 words, with distribution ranging from 0 to 268 words. The median was 10 words, whereas the mode was 1 word, produced by 135 kanji. The mean of kanji lexical productivity on the left-hand side was 19.63 words, with a standard deviation of 26.80 words. The distribution was positively skewed (skewness = 3.08) with a high peak (kurtosis = 13.67).

Using the same counting procedure, type frequencies of two-kanji compound words on the right-hand side are provided in Row 12. As is shown in Table 4, a large number of 1,365 kanji (70.18% of the 1,945 kanji) produced

Table 3
Distribution of Kanji Lexical Productivity on the Left-Hand Side

Range of Kanji Productivity	Number of Kanji	Percentage	Accumulative Percentage
0–9	931	47.87	47.87
10–19	420	21.59	69.46
20–29	202	10.39	79.85
30–39	122	6.27	86.12
40–49	82	4.22	90.33
50–59	47	2.42	92.75
60–69	41	2.11	94.86
70–79	20	1.03	95.89
80–89	17	0.87	96.76
90	10	0.51	97.28
100	12	0.62	97.89
110	10	0.51	98.41
120	10	0.51	98.92

Note—Figures beyond an accumulative percentage of 99% were excluded.

0–19 two-kanji compound words. The highest and second highest kanji lexical productivity differed between the left-hand and the right-hand sides: the kanji 大 (/dai/ in On-reading and /oR/ in Kun-reading, "big") showed the highest kanji lexical productivity on the right-hand side, producing 399 words. The second highest was 一 (/itJi/ in On-reading and /hito(tu)/ in Kun-reading, "one"), having 350 words. The distribution ranged from 0 to 399 words. The median was 10 words, whereas the mode was 3 words. The mean of kanji lexical productivity on the right-hand side was 19.69 words, with a standard deviation of 28.19 words. The distribution was positively skewed (skewness = 4.59) with a high peak (kurtosis = 37.00).

The total kanji lexical productivity for each kanji was produced by adding type frequencies of both the left- and the right-hand sides. For example, the kanji 山 (/saN/ in On-reading, /Jama/ in Kun-reading, "mountains") produced 118 different two-kanji compound words by changing the kanji on the left-hand side, such as 鉱山 (/koR zaN/, "mine"), 下山 (/ge zaN/, "to descend a mountain"), and 小山 (/ko jama/, "hill"), whereas the same kanji produced 163 different words on the right-hand side, such as 山河 (/saN ga/ "mountains and rivers"), 山城 (/jama siro/, "a mountain castle"), and 山頂 (/saN tjoR/, "mountain top"). Thus, the total of kanji lexical productivity for the kanji 山 becomes 281 words (118 plus 163). As is shown in Table 5, a large number of 858 kanji (44.11% of 1,945 kanji) produced 0–19 two-kanji compound words. The kanji 大 exhibited the highest kanji lexical productivity on both the left- and the right-hand sides together, producing 469 words (70 on the left-hand side and 399 on the right-hand side). The second highest was 人 (268 on the left-hand side and 137 on the right-hand side), creating 405 words. The distribution ranged from 0 to 469 words. The median of kanji was 23 words, whereas the mode was 11 words. The mean of kanji lexical productivity was 39.32 words, with a standard deviation of 47.70 words. The distribution was positively skewed (skewness = 3.10) with a high peak (kurtosis = 13.71).

Pearson's correlations revealed that kanji lexical productivity on the left-hand side for the 1,945 basic kanji was reasonably high at .505 ($p < .01$), with one kanji on the right-hand side. This suggests that numbers of kanji lexical productivity have some relation between the left- and the right-hand sides, but these are not always equally proportional on both sides. The figures of the total kanji lexical productivity (n = 1,945) were highly correlated at .644 ($p < .01$) with printed frequency in 2000, at .668 ($p < .01$) with printed frequency in 1998 (M = 8,590.00, SD = 18,042.50), at .673 ($p < .01$) with CD-ROM frequency in 1998 (M = 12,484.16, SD = 26,089.57), and at .674 ($p < .01$) with printed frequency in 1976 (M = 0.51, SD = 1.09, per thousand appearance of total kanji frequency). Although it is rather natural to expect a high correlation between single kanji frequencies and their lexical productivities, the correlations were lower than .700. Interestingly, the same figures of kanji lexical productivity were highly correlated at .639 ($p < .01$) with the levels of the Japanese Language Proficiency Test for international students (M = 1.69, SD = 0.80; for details, see Aruku Nihongo Shuppan Henshuubu, 2000; Japan Foundation and Association of International Education, 1996; Tamaoka et al., 2002), and at −.666 ($p < .01$) with school grades in which each kanji is taught in Japanese schools (M = 5.33, SD = 1.96). Stroke numbers of kanji (M = 10.34, SD = 3.76) showed a modest correlation at −.287 ($p < .01$) with kanji lexical productivity.

Table 4
Distribution of Kanji Lexical Productivity on the Right-Hand Side

Range of Kanji Productivity	Number of Kanji	Percentage	Accumulative Percentage
0–19	1,365	70.18	70.18
20–39	318	16.35	86.53
40–59	138	7.10	93.62
60–79	57	2.93	96.56
80–99	21	1.08	97.63
100–119	16	0.82	98.46

Note—Figures beyond an accumulative percentage of 99% were excluded.

Table 5
Distribution of Kanji Lexical Productivity on the Left- and Right-Hand Sides

Range of Kanji Productivity	Number of Kanji	Percentage	Accumulative Percentage
0–19	858	44.11	44.11
20–39	465	23.91	68.02
40–59	261	13.42	81.44
60–79	127	6.53	87.97
80–99	68	3.50	91.47
100–119	50	2.57	94.04
120–139	32	1.65	95.68
140–159	23	1.18	96.86
160–179	15	0.77	97.63
180–199	8	0.41	98.05
200–219	7	0.36	98.41
220–239	8	0.41	98.82

Note—Figures beyond an accumulative percentage of 99% were excluded.

Accumulative Kanji Lexical Productivity on the Left- and the Right-Hand Sides in Rows 14–16

Kanji lexical productivity in Rows 11–13 are simply a count of two-kanji compound words produced by a single kanji with no consideration of word frequency. Accumulative word frequency of all words together is considered to be more accurate in indicating the magnitude of kanji lexical productivity than is a simple count of each produced word as "1" (e.g., Grainger et al., 1989). Thus, accumulative kanji lexical productivity is calculated by adding all the frequencies of occurrence for words in print provided by Amano and Kondo (2000). Since two-kanji compound words can be formed by combining kanji on the left- or the right-hand side of a single targeted kanji, figures of accumulative kanji lexical productivity are provided for the 1,945 basic kanji on the left-hand side (Row 14), the right-hand side (Row 15), and both sides together (Row 16).

As is shown in Table 6, a great majority of the 1,945 kanji (1,483, or 76.25%) produced words appearing 0 to 19,999 times on the left-hand side. Among them, the kanji 長 showed the highest accumulative kanji lexical productivity, counted 465,542 times for 102 compound words. The second highest was 業, counted 448,611 times for 123 words. The distribution ranged from 0 to 465,542 times. The median was 2,893 times, whereas the mode was 0 times. The mean was 22,247.87 times, with a very large standard deviation of 49,858.76 times. The distribution was positively skewed (skewness = 4.20) with a very high peak (kurtosis = 22.39).

Similarly, as is shown in Table 7, a large majority of the 1,945 kanji (1,471, or 75.63%) produced words appearing 0 to 19,999 times on the right-hand side. The kanji 政 showed the highest accumulative kanji lexical productivity, counted 566,590 times for only 37 words. Since the lexical database of Amano and Kondo (2000) counted word frequencies in the *Asahi* newspaper, some specific words such as 政治 (/sei zi/, "politics") and 政界 (/sei kai/, "political world"), must have been used frequently in various articles. The second highest was 会, counted 517,963 times for 43 words. Again, this kanji produces some high-frequency words, such as 会議 (/kai gi/, "meeting" or "congress") and 会合 (/kai goR/, "meeting" or "assembly"), which renders this kanji the second highest in accumulative kanji lexical productivity on the right-hand side. The distribution ranged from 0 to 566,590 times. The median was 3,714 times, whereas the mode was 0 times. The mean was 22,246.91 times, with a very large standard deviation of 50,281.75 times. As with the basic kanji on the left-hand side, the same values on the right-hand side showed similar distribution, positive skewness (skewness = 4.68) and a very high peak (kurtosis = 30.00).

A distribution of a total of accumulative kanji lexical productivity for the 1,945 basic kanji on the left- and right-hand sides together is reported in Table 8. A large majority of the 1,945 kanji (1,417, or 72.85%) produced words appearing 0 to 39,999 times. The kanji 会 showed the highest accumulative kanji lexical productivity, counted 920,283 times for 128 compound words. The second highest was 国, counted 811,848 times for 248 words. The distribution ranged from 0 to 920,283 times. The median was 10,749 times, whereas the mode was 0 times. The mean was 44,494.78 times, with a very large standard deviation of 85,666.47 times. The distribution was positively skewed (skewness = 3.93) with a high peak (kurtosis = 21.00).

Pearson's correlations (n = 1,945) for accumulative kanji lexical productivity on the left-hand side, at .464 ($p < .01$), were considerably high with ones on the right-hand side. The figures for total accumulative kanji lexical productivity were highly correlated at .854 ($p < .01$) with printed frequency in 2000, at .652 ($p < .01$) with printed frequency in 1998, at .645 ($p < .01$) with CD-ROM frequency in 1998, and at .581 ($p < .01$) with printed frequency in 1976. Again, similar to correlations of kanji lexical productivity, accumulative figures had high correlations between single kanji frequencies and their lexical productivities. The same figures of the accumulative kanji lexical productivity were correlated at .413 ($p < .01$) with the levels of the Japanese Language Proficiency Test for international students and at −.403

Table 6
Distribution of Accumulative Kanji Lexical Productivity on the Left-Hand Side

Range of Kanji Productivity	Number of Kanji	Percentage	Accumulative Percentage
0–19,999	1,483	76.25	76.25
20,000–39,999	165	8.48	84.73
40,000–59,999	88	4.52	89.25
60,000–79,999	49	2.52	91.77
80,000–99,999	44	2.26	94.04
100,000–119,999	32	1.65	95.68
120,000–139,999	15	0.77	96.45
140,000–159,999	13	0.67	97.12
160,000–179,999	11	0.57	97.69
180,000–199,999	6	0.31	97.99
200,000–219,999	9	0.46	98.46
220,000–239,999	5	0.26	98.71
240,000–259,999	5	0.26	98.97

Note—Figures beyond an accumulative percentage of 99% were excluded.

Table 7
Distribution of Accumulative Kanji Lexical Productivity on the Right-Hand Side

Range of Kanji Productivity	Number of Kanji	Percentage	Accumulative Percentage
0–19,999	1,471	75.63	75.63
20,000–39,999	190	9.77	85.40
40,000–59,999	78	4.01	89.41
60,000–79,999	61	3.14	92.54
80,000–99,999	29	1.49	94.04
100,000–119,999	26	1.34	95.37
120,000–139,999	18	0.93	96.30
140,000–159,999	16	0.82	97.12
160,000–179,999	15	0.77	97.89
180,000–199,999	6	0.31	98.20
200,000–219,999	10	0.51	98.71
220,000–239,999	5	0.26	98.97

Note—Figures beyond an accumulative percentage of 99% were excluded.

Table 8
Distribution of Accumulative Kanji Lexical Productivity on Both the Left- and the Right-Hand Sides

Range of Kanji Productivity	Number of Kanji	Percentage	Accumulative Percentage
0–39,999	1,417	72.85	72.85
40,000–79,999	210	10.80	83.65
80,000–119,999	118	6.07	89.72
120,000–159,999	59	3.03	92.75
160,000–199,999	42	2.16	94.91
200,000–239,999	21	1.08	95.99
240,000–279,999	19	0.98	96.97
280,000–319,999	15	0.77	97.74
320,000–359,999	11	0.57	98.30
360,000–399,999	10	0.51	98.82

Note—Figures beyond an accumulative percentage of 99% were excluded.

($p < .01$) with grades in which each kanji is taught in Japanese schools. Stroke numbers of kanji showed a low correlation at $-.131$ ($p < .01$) with accumulative kanji lexical productivity.

Symmetry in Row 17

Each kanji creates various two-kanji compound words by changing another kanji on the left-hand or right-hand side, as explained in the kanji lexical productivity (Rows 11–13). Symmetry indicates an equal tendency of kanji lexical productivity (type frequency) on both the left- and the right-hand sides. For example, the kanji, 紺 (/KON/ in On-reading, "navy blue" and no sound in Kun-reading) produces 10 different two-kanji compound words, 2 words on the left-hand side and 8 words on the right-hand side. An asymptotic test for symmetry of individual kanji can be performed as follows: Let the number of left-hand side compounds be n_L, and those of the right-hand side n_R and $n_L + n_R = n$. Then, under the hypothesis of equality of both sides, the expected value is $n/2$. The asymptotic chi-square criterion is

$$X^2 = \frac{\left(n_L - \frac{n}{2}\right)^2}{\frac{n}{2}} + \frac{\left(n_R - \frac{n}{2}\right)^2}{\frac{n}{2}}$$

$$= \frac{\left(n_L - \frac{n}{2}\right)^2}{\frac{n}{2}} + \frac{\left(n - n_L - \frac{n}{2}\right)^2}{\frac{n}{2}}$$

$$= \frac{\left(n_L - n_R\right)^2}{n_L + n_R},$$

which is distributed as a chi-square with one degree of freedom. At the $\alpha = .05$ level, this must be greater than 3.84 in order to be significant. For example, lexical productivity for the kanji 紺 is $n_L = 2$ and $n = 10$, yielding

$$X^2 = \frac{(2-8)^2}{2+8} = 3.6,$$

which, at less than 3.84, is not significant. Therefore, the kanji 紺 is not symmetric, but asymmetric.

There were 227 kanji (11.67% of the total) with fewer than five words having the total kanji lexical productivity putting both the left- and the right-hand sides together; therefore, the total of kanji lexical productivity is too small to test their symmetry (recoded as "." in the kanji database). Thus, the rest of the 1,718 kanji were tested for symmetry. Excluding these 227 kanji, 902 kanji (52.50% out of 1,718 kanji) were judged to be symmetric, represented by "S." When left-hand side productivity was greater than that of the right-hand side, a kanji was judged as progressively asymmetric, indicated by "P." Four hundred three kanji (23.46% out of 1,718 kanji) fell into this category. When the right-hand side productivity was greater than the left-hand side, a kanji was judged as regressively asymmetric, represented by "R." Four hundred thirteen kanji (24.04% out of 1,718 kanji) were counted in this category. Simply looking at the percentages, a set of the whole basic kanji seems to display a symmetric pattern of lexical productivity. However, there is no tendency, on the whole, for basic kanji to produce compound words to the same extent on the left-hand or the right-hand side; in fact, this was highly asymmetric (for details, see Tamaoka & Altmann, 2004).

Table 9
Kanji Symmetry of Left- and Right-Hand Lexical Productivity

Type of Symmetry	Number of Kanji	Percentage	Accumulative Percentage
1	403	23.46	23.46
2	902	52.50	75.96
3	413	24.04	100.00
Total	1,718	100.00	

Note—Kanji with lower than 5 lexical productivity on both the left- and the right-hand sides are excluded for symmetry testing. Thus, the total number of kanji is 1,718.

Kanji Entropy on the Left-Hand Side in Row 18 and the Right-Hand Side in Row 19

For kanji, entropy refers to how randomly each kanji produces a two-kanji compound, and it is calculated using the formula

$$H = -\sum p_j \log_2 p_j.$$

Kanji entropy is calculated on the basis of its lexical productivity. For example, the kanji 亜 (/a/ in On-reading and /tu(gu)/ in Kun-reading, "rank next" or "come after") produces three words of 興亜 (/KOA a/, "Asia development," 61 printed frequency), 東亜 (/TOA a/, "East Asia," 706 printed frequency) and 白亜 (/haku a/, "chalk," 90 printed frequency), combining with three different kanji on the left-hand side. The p in the formula stands for the probability that a specific word will appear among all the compound words combined with multiple kanji on the left-hand side of the kanji 亜. In the case of the word 興亜, p is .0712, as calculated by dividing 61 by a total word printed frequency of 858. The formula $\log_2 p_j$ for the word 興亜 is simply calculated as $\log_2 .071 = -3.8124$. Likewise, $p_j \log_2 p_j$ for 東亜 is $-.2304$ (the result of $.8238 \times -.2796$). In the same manner, the word 白亜 is $-.3414$. The kanji entropy of 亜 on the left-hand side is finally determined as .8432 by adding the scores of $-.2714$, $-.2304$, and $-.3414$ and dividing by -1. Entropy values for the basic kanji on the right-hand side were calculated in the same way.

The distribution of kanji entropy on the left-hand side is reported in Table 10. Entropies of 377 kanji (19.38% of 1,945 kanji) were calculated from 0.00 to 0.29. Since 243 kanji had no or only one word on the left-hand side produced by a target kanji, their entropies all became 0.00. The kanji 水 (/sui/ in On-reading and /mizu/ in Kun-reading, "water"), which produced 162 words with a total word frequency of 32,347 times on the left-hand side, showed the highest kanji entropy of 5.568 (the kanji database reports kanji entropy to three decimal points). The second highest entropy was 人, calculated as 5.359, by

Table 10
Distribution of Kanji Entropy on the Left-Hand Side

Range of Kanji Entropy	Number of Kanji	Percentage	Accumulative Percentage
0.00–0.29	377	19.38	19.38
0.30–0.59	109	5.60	24.99
0.60–0.89	121	6.22	31.21
0.90–1.19	176	9.05	40.26
1.20–1.49	155	7.97	48.23
1.50–1.79	154	7.92	56.14
1.80–2.09	157	8.07	64.22
2.10–2.39	147	7.56	71.77
2.40–2.69	153	7.87	79.64
2.70–2.99	135	6.94	86.58
3.00–3.29	97	4.99	91.57
3.30–3.59	63	3.24	94.81
3.60–3.89	40	2.06	96.86
3.90–4.19	33	1.70	98.56
4.20–4.49	13	0.67	99.23
4.50–4.79	9	0.46	99.69
4.80–5.09	4	0.21	99.90
5.10–5.39	1	0.05	99.95
5.40–	1	0.05	100.00
Total number of kanji			1,945
Total kanji entropy			3,146.79
Maximum kanji entropy			5.57
Minimum kanji entropy			0.00
Median			1.56
Mode			0.00
Mean			1.62
Standard deviation			1.18
Skewness			0.31
Kurtosis			−0.71

Table 11
Distribution of Kanji Entropy on the Right-Hand Side

Range of Kanji Entropy	Number of Kanji	Percentage	Accumulative Percentage
0.00–0.29	283	14.55	14.55
0.30–0.59	121	6.22	20.77
0.60–0.89	146	7.51	28.28
0.90–1.19	205	10.54	38.82
1.20–1.49	162	8.33	47.15
1.50–1.79	175	9.00	56.14
1.80–2.09	173	8.89	65.04
2.10–2.39	159	8.17	73.21
2.40–2.69	152	7.81	81.03
2.70–2.99	109	5.60	86.63
3.00–3.29	81	4.16	90.80
3.30–3.59	67	3.44	94.24
3.60–3.89	52	2.67	96.92
3.90–4.19	33	1.70	98.61
4.20–4.49	14	0.72	99.33
4.50–4.79	7	0.36	99.69
4.80–5.09	2	0.10	99.79
5.10–5.39	3	0.15	99.95
5.40–5.69	0	0.00	99.95
5.70–	1	0.05	100.00
Total number of kanji			1,945
Total kanji entropy			3,229.54
Maximum kanji entropy			5.79
Minimum kanji entropy			0.00
Median			1.59
Mode			0.00
Mean			1.66
Standard deviation			1.14
Skewness			0.36
Kurtosis			−0.54

producing 268 words with a total word frequency of 402,601 times. Even though entropies of these top two kanji showed similar figures, numbers of kanji lexical productivity and their total word frequencies were distinctly different. Thus, kanji entropy is not exactly representative of kanji lexical productivity. As is shown in Table 10, the distribution ranged from 0.00 to 5.57. The median of kanji entropy was 1.56, whereas the mode was 0.00. The mean of kanji entropy on the left-hand side was 1.62, with a standard deviation of 1.18. The distribution showed a slight positive skew (skewness = 0.31) with a relatively low peak (kurtosis = −0.71).

The distribution of kanji entropy on the right-hand side is reported in Table 11. Similar to those on the left-hand side, entropies of 283 kanji (14.55% of 1,945 kanji) were calculated from 0.00 to 0.29. Among them, 163 kanji showed an entropy of 0.00. The kanji 一, which produced 350 words with a total word frequency of 466,757 times on the right-hand side, showed the highest kanji entropy of 5.79, whereas the second highest entropy was exhibited by 同 (/doR/ in On-reading and /ona(zi)/ in Kun-reading, "same") calculated as 5.18, by producing 148 words with a total word frequency of 279,528 times. The distribution ranged from 0.00 to 5.79. The median of kanji entropy was 1.59, and the mode was 0.00. The mean of kanji entropy on the right-hand side was 1.66, with a standard deviation of 1.14. The distribution showed a slight positive skew (skewness = 0.36) with a low peak (kurtosis = −0.54).

Pearson's correlations (n = 1,945) indicated that figures for kanji entropy on the left-hand side were moderate at .390 ($p < .01$), as compared with those on the right-hand side. Kanji entropy on the left-hand side showed moderate correlations of .360 ($p < .01$) with printed frequency in 2000, .316 ($p < .01$) with printed frequency in 1998, .319 ($p < .01$) with CD-ROM frequency in 1998, and .318 ($p < .01$) with printed frequency in 1976. Similarly, kanji entropy on the right-hand side indicated moderate correlations of .346 ($p < .01$) with printed frequency in 2000, .379 ($p < .01$) with printed frequency in 1998, .383 ($p < .01$) with CD-ROM frequency in 1998, and .393 ($p < .01$) with printed frequency in 1976. Overall, kanji entropy on either side did not strongly correlate with kanji printed frequency. On the other hand, as for kanji lexical productivity, kanji entropy on the left-hand side showed a high correlation of .719 ($p < .01$) with kanji lexical productivity on the left-hand side. Likewise, kanji entropy on the right-hand side was highly correlated at .637 ($p < .01$) with kanji lexical productivity on the right-hand side. However, accumulative kanji lexical productivity showed only moderate correlations with kanji entropy, at .380 ($p < .01$) on the left-hand side and .358 ($p < .01$) on the right-hand side. Correlations between kanji entropy and the levels of the Japanese Language Proficiency Test proved reasonably high at .410 ($p < .01$) on the left-hand side and .457 ($p < .01$) on the right-hand side. School grades in which each kanji is taught in Japanese schools revealed slightly higher correlations than those for the Japanese

Language Proficiency Test, at $-.499$ ($p < .01$) on the left-hand side and $-.518$ ($p < .01$) on the right-hand side.

Kanji Redundancy on the Left-Hand Side in Row 20 and the Right-Hand Side in Row 21

In addition to entropy, another well-known mathematical concept by Shannon (1948) is *redundancy*, which refers to the degree of superfluousness. For kanji, redundancy implies frequency bias in appearance, indicating, for instance, repeated use of compound words in the same lexical corpus of Amano and Kondo (2000). Redundancy is determined using the formula

$$R = (1 - H/H_{max}) \times 100\ (\%).$$

H refers to entropy, whereas H_{max} indicates maximum entropy. Maximum entropy for 亜 implies that any kanji can be combined with the specific kanji with the same probability, as produced by the formula

$$H_{max} = \log_2 J.$$

As with the aforementioned example of 亜 on the right-hand side, H_{max} is calculated to be 1.585, using logarithm $\log_2$ of type frequency $J = 3$. This figure shows the entropy for three different kanji on the left-hand side to be equal (i.e., the same token frequency) when combined with 亜. Redundancy is now calculated as a percentage; the result of the figure for entropy (0.8432) divided by the maximum entropy (1.5850) is subtracted from 1 (i.e., 1 − 0.8432/1.5850), and then its figure (0.47) is multiplied by 100 to render it a percentage (46.8%). Thus, the redundancy with 亜 on the right-hand side equals 46.8%. However, figures of kanji redundancy in the kanji database were recorded to three decimal points—that is, 0.468. Kanji redundancies on the left-hand side were calculated in the same manner. All kanji redundancy values were produced using this procedure, except 243 kanji on the left-hand side and 165 kanji on the right with 0.00 entropy (redundancy cannot be calculated with 0.00 entropy).

As is shown in Table 12, the distribution of kanji redundancy on the left-hand side (n = 1,702) seems to have a smooth bell-shaped curve. The distribution showed a slightly positive skew (skewness = 0.34) with a low peak (kurtosis = −0.20). Both the kanji 模 (/mo/ in On-reading and /katado(ru)/ in Kun-reading, "imitate") and 摘 (/teki/ in On-reading and /tuma(mu)/ in Kun-reading, "pick") were calculated to show the highest kanji redundancy of 99.9%. Both kanji had few words produced on the left-hand side, whereas their word frequencies were high: 模 had only two words with a frequency of 30,641 times (entropy = 0.001), whereas 摘 had three words with a frequency of 62,952 times (entropy = 0.002). The distribution ranged from 0% to 99.95%. The median was 47.98%, whereas there were no equal redundancy figures (nil modes). The mean was 50.87%, and the standard deviation was 21.27%.

As with the left-hand side, the distribution of kanji redundancy on the right-hand side (n = 1,780) resulted in a similar smooth curve, shown in Table 13. The distribution showed a slightly positive skew (skewness = 0.34) with a low peak (kurtosis = −0.14). Four kanji of 逮 (/tai/ in On-reading and no Kun-reading, "chase"), 誕 (/tan/ in On-reading and no Kun-reading, "be born"), 撮 (/satu/ in On-reading, /to(ru)/ in Kun-reading, "pinch"), and 距 (/kjo/ in On-reading, /heda(taru)/ in Kun-reading, "be distance") showed the highest kanji redundancy of 99.9%. These kanji produced only two compound words by changing kanji on the right-hand side but had a high frequency of 42,796 times for 逮, 14,060 times for 誕, 12,866 times for 撮, and 10,234 times for 距. Entropy for these four kanji was 0.001. The distribution ranged from 0% to 99.89% (or 99.9%). The median was 45.58%, whereas the mode was 8.17%. The mean was 51.41% with a standard deviation of 20.13%.

Pearson's correlations indicated that figures of kanji redundancy on the left-hand side (n = 1,588) had a very weak correlation of .039 (n.s.) with ones on the right-hand side. Kanji redundancy on the left-hand side (n = 1,702) showed very weak correlations of .034 (n.s.) with printed frequency in 2000, .016 (n.s.) with printed frequency in 1998, .016 (n.s.) with CD-ROM frequency in 1998, and .004 (n.s.) with printed frequency in 1976. Similarly, kanji redundancy on the right-hand side (n = 1,780) indicated very weak correlations of $-.013$ (n.s.) with printed frequency in 2000, $-.035$ (n.s.) with printed frequency in 1998, $-.036$ (n.s.) with CD-ROM frequency in 1998, and $-.047$ (n.s.) with printed frequency in 1976.

Table 12
Distribution of Kanji Redundancy on the Left-Hand Side

Range of Kanji Redundancy (%)	Number of Kanji	Percentage	Accumulative Percentage
0.00–4.99	20	1.18	1.18
5.00–9.99	23	1.35	2.53
10.00–14.99	23	1.35	3.88
15.00–19.99	24	1.41	5.29
20.00–24.99	50	2.94	8.23
25.00–29.99	91	5.35	13.57
30.00–34.99	141	8.28	21.86
35.00–39.99	178	10.46	32.31
40.00–44.99	198	11.63	43.95
45.00–49.99	185	10.87	54.82
50.00–54.99	148	8.70	63.51
55.00–59.99	119	6.99	70.51
60.00–64.99	98	5.76	76.26
65.00–69.99	70	4.11	80.38
70.00–74.99	76	4.47	84.84
75.00–79.99	62	3.64	88.48
80.00–84.99	49	2.88	91.36
85.00–89.99	38	2.23	93.60
90.00–94.99	47	2.76	96.36
95.00–100.00	62	3.64	100.00
Total number of kanji			1,702
Total kanji redundancy			86,575.91
Maximum kanji redundancy			99.95
Minimum kanji redundancy			0.00
Median			47.9
Mode			0.00
Mean			50.87
Standard deviation			21.27
Skewness			0.34
Kurtosis			−0.20

Table 13
Distribution of Kanji Redundancy on the Right-Hand Side

Range of Kanji Redundancy (%)	Number of Kanji	Percentage	Accumulative Percentage
0.00–4.99	15	0.84	0.84
5.00–9.99	15	0.84	1.69
10.00–14.99	19	1.07	2.75
15.00–19.99	21	1.18	3.93
20.00–24.99	47	2.64	6.57
25.00–29.99	85	4.78	11.35
30.00–34.99	155	8.71	20.06
35.00–39.99	186	10.45	30.51
40.00–44.99	223	12.53	43.03
45.00–49.99	170	9.55	52.58
50.00–54.99	175	9.83	62.42
55.00–59.99	140	7.87	70.28
60.00–64.99	103	5.79	76.07
65.00–69.99	91	5.11	81.18
70.00–74.99	88	4.94	86.12
75.00–79.99	62	3.48	89.61
80.00–84.99	59	3.31	92.92
85.00–89.99	33	1.85	94.78
90.00–94.99	43	2.42	97.19
95.00–100.00	50	2.81	100.00
Total number of kanji			1,780
Total kanji redundancy			91,508.81
Maximum kanji redundancy			99.89
Minimum kanji redundancy			0.00
Median			48.58
Mode			8.17
Mean			51.41
Standard deviation			20.13
Skewness			0.34
Kurtosis			−0.14

Overall, kanji redundancy on either side correlated very weakly with kanji printed frequency. Similarly, kanji redundancy on the left-hand side (n = 1,702) showed weak correlations of .123 ($p < .01$) with kanji lexical productivity on the left-hand side, with kanji redundancy on the right-hand side (n = 1,780), and with kanji lexical productivity on the right-hand side ($r = -.146$, $p < .01$). The accumulative kanji lexical productivity showed even weaker correlations of .033 (n.s.) with kanji redundancy on the left-hand side (n = 1,702) and −.009 (n.s.) on the right-hand side (n = 1,780). Correlations between kanji redundancy and the levels of the Japanese Language Proficiency Test were weak at −.016 (n.s.) on the left-hand side (n = 1,695) and −.050 ($p < .05$) on the right-hand side (n = 1,773). School grades in which each kanji is taught correlated weakly with kanji redundancy at .028 (n.s.) on the left-hand side (n = 1,702) and .070 ($p < .01$) on the right (n = 1,780). Despite all of these low correlations, correlations between kanji redundancy and kanji entropy were highly correlated at −.626 ($p < .01$) on the left-hand side and −.634 ($p < .01$) on the right. Although these figures for redundancy and entropy remain relatively highly correlated, they do not have a perfect negative correlation of −1.00. In other words, these two indexes are seen to measure different phenomena.

Numbers of Meanings for On-Readings in Row 26 and Kun-Readings in Row 1

Kanji pronunciation can be divided into two types: On-readings derived from the original Chinese pronunciation, and Kun-readings originating from the Japanese pronunciation (for details, see Kabashima, 1989; Kess & Miyamoto, 1999; Takashima, 2001; Tamaoka, 1991). Native Japanese speakers are able to judge On-readings and Kun-readings at a relatively high probability rate (Tamaoka, 2003) when they listen to kanji pronunciations. Usage of On- and Kun-readings relates to word origin. For example, the kanji 空 ("sky") is pronounced as /kuR/ in On-reading. On-reading is typically used for kanji compound words, such as 空気 (/kuR ki/, "air"), 空中 (/kuR tjuR/, "in the air"), 空港 (/kuR koR/, "airport"), and 空想 (/kuR soR/, "fantasy"). On the other hand, the kanji 空 /sora/, in Kun-reading, is a single word originating from the traditional Japanese vocabulary called *wago*. A majority of kanji are used as single, free-standing lexical units originating from *wago* and are read in Kun-reading. Thus, native Japanese speakers likely use Kun-reading for single free-standing nouns presented in a single kanji. *Wago* is also written using two kanji, such as 空耳 (/sora mimi/, "mishearing") and 空言 (/sora goto/, "falsehood"). As such, On- and Kun-readings are used distinctly for different words: On-readings for *kango* (traditional Chinese words) and Kun-readings for *wago* (traditional Japanese words). A number of meanings were counted using semantic headings in each kanji according to a basic Japanese kanji dictionary (Kamata, 1991). Since meanings occasionally differ between On-readings and Kun-readings, numbers of meanings were counted separately for On-reading in Row 26 and for Kun-readings in Row 31.

Numbers of the meanings for On-readings were reported in Table 14. A large number of 1,187 kanji (61.03% of the 1,945 kanji) had only one (639 kanji) or two meanings (548 kanji) for On-reading. The distribution had a positive skew (skewness = 1.93) with a slightly higher peak (kurtosis = 5.95). Two kanji, 分(/huN/ in On-reading and /wake(ru)/ in Kun-reading, "divide") and 発 (/hatu/ in On-reading and /ta(tu)/ in Kun-reading, "leave") had the highest number of 14 different meanings for On-

Table 14
Distribution of the Number of Meanings for On-Readings

Number of Meanings in On-	Number of Kanji	Percentage	Accumulative Percentage
0	40	2.06	2.06
1	639	32.85	34.91
2	548	28.17	63.08
3	340	17.48	80.57
4	176	9.05	89.61
5	89	4.58	94.19
6	55	2.83	97.02
7	24	1.23	98.25
8	10	0.51	98.77

Note—Figures beyond an accumulative percentage of 99% were excluded.

Table 15
Distribution of the Number of Meanings for Kun-Readings

Number of Meanings in Kun-	Number of Kanji	Percentage	Accumulative Percentage
0	737	37.89	37.89
1	291	14.96	52.85
2	279	14.34	67.20
3	191	9.82	77.02
4	127	6.53	83.55
5	80	4.11	87.66

Note—Figures beyond an accumulative percentage of 99% were excluded.

readings. Since 40 kanji had no On-readings, these kanji had zero meanings for On-readings. The distribution ranged from 0 to 14. The median was 2, and the mode was 1. The mean was 2.43, with a standard deviation of 1.73.

Numbers of the meanings for Kun-readings are reported in Table 15. In that 737 kanji (37.89% of the 1,945 kanji) had no Kun-readings, they had zero meanings. The distribution was positively skewed (skewness = 3.93) with a very high peak (kurtosis = 25.95). The kanji 立 (/ritu/ in On-reading and /ta(tu)/ in Kun-reading, "stand") had the highest of 41 different meanings for Kun-readings. The distribution ranged from 0 to 41 meanings. The median was 1, and the mode was 0. The mean was 2.37, with a standard deviation of 3.61.

Pearson's correlations (n = 1,945) between the numbers of meanings for On- and Kun-readings was rather low, at .295 ($p < .01$). Since some kanji have zero meanings for either On-readings (40 kanji) or Kun-readings (737 kanji), after excluding these kanji, the correlation (n = 1,177), at .365 ($p < .01$) was still lower than had been expected. These correlation results suggest that meanings for On-readings do not necessarily share the same meanings for Kun-readings and vice versa. The numbers of meanings for On-readings (n = 1,945) correlated moderately with kanji entropy at .380 ($p < .01$) on the left-hand side and at .439 ($p < .01$) on the right-hand side, whereas the numbers of meanings for Kun-readings (n = 1,945) showed much weaker correlations with kanji entropy at .220 ($p < .01$) on the left-hand side and at .292 ($p < .01$) on the right-hand side. Interestingly, the same figures had almost no correlations with kanji redundancy. The numbers of meanings for On-readings were correlated at .005 (n.s.) on the left-hand side (n = 1,702) and at −.068 ($p < .01$) on the right-hand side (n = 1,780), whereas the numbers of meanings for Kun-readings were correlated at .010 (n.s.) on the left-hand side (n = 1,702) and −.038 (n.s.) on the right (n = 1,780).

The meanings for On-readings (n = 1,945) showed a considerably high correlation of .419 ($p < .01$) with printed frequency in 2000, .371 ($p < .01$) with printed frequency in 1998, .371 ($p < .01$) with CD-ROM frequency in 1998, and .379 ($p < .01$) with printed frequency in 1976. The meanings for Kun-readings were lower than those for On-readings (n = 1,945), being correlated at .228 ($p < .01$) with printed frequency in 2000, .262 ($p < .01$) with printed frequency in 1998, .259 ($p < .01$) with CD-ROM frequency in 1998, and .254 ($p < .01$) with printed frequency in 1976. Overall, the numbers of meanings for either On- or Kun-readings remained moderately correlated with kanji printed frequency.

On the other hand, the numbers of meanings for On-readings indicated relatively high correlations with kanji lexical productivity (n = 1,945) at .449 ($p < .01$) on the left-hand side and at .506 ($p < .01$) on the right-hand side. The correlations were moderately high but *weaker* for Kun-readings, with the same figures of kanji lexical productivity (n = 1,945) at .308 ($p < .01$) on the left-hand side and at .362 ($p < .01$) on the right-hand side. A similar trend in correlations was observed with accumulative kanji lexical productivity. Since a large majority of two-kanji compound words were pronounced in On-readings, stronger correlations were seen for On-readings than for Kun-readings. Correlations between the numbers of meanings and the levels of the Japanese Language Proficiency Test (n = 1,926) were shown to be moderately high at .356 ($p < .01$) for On-readings and .324 ($p < .01$) for Kun-readings. School grades in which each kanji is taught (n = 1,945) revealed slightly higher correlations than those for the Japanese Language Proficiency Test, being at −.415 ($p < .01$) for On-readings and lower at −.301 ($p < .01$) for Kun-readings. Generally speaking, the numbers of meanings for On-readings show higher correlations with various figures than do those for Kun-readings.

REFERENCES

AMANO, N., & KONDO, K. (2000). *Nihongo-no goi tokusei* [Lexical properties of Japanese]. Tokyo: Sanseido.

ANDREWS, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 802-814.

ARUKU NIHONGO SHUPPAN HENSHUUBU [Department of Japanese Publishing, Aruku] (2000). *Nihongo nooryoku shiken kanji handobukku* [A kanji handbook for the Japanese Language Proficiency Test]. Tokyo: Aruku.

COLTHEART, M., DAVELAAR, E., JONASSON, J. T., & BESNER, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). New York: Academic Press.

GRAINGER, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory & Language*, **29**, 228-244.

GRAINGER, J. (1992). Orthographic neighborhoods and visual word recognition. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology and meaning* (pp. 131-166). Amsterdam: North-Holland.

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.

GRAINGER, J., O'REGAN, J. K., JACOBS, A. M., & SEGUI, J. (1992). Neighborhood frequency effects and letter visibility in visual word recognition. *Perception & Psychophysics*, **51**, 49-56.

GRAINGER, J., & SEGUI, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception & Psychophysics*, **47**, 191-198.

HAYASHI, S. (1987). *Kanji, goi, bunshoo-no kenkyuu-e* [Toward the study of kanji, words and sentences]. Tokyo: Meiji Shoin.

HORI, J. (1979). *Entoropii towa nanika* [What is entropy?] Tokyo: Koodansha Blue Books.

JAPAN FOUNDATION AND ASSOCIATION OF INTERNATIONAL EDUCATION (1996). *Nihongo nooryoku shiken: Shutsudai kijun* [The Japanese

Language Proficiency Test: Test content specifications]. Tokyo: Bonjinsha.

KABASHIMA, T. (1989). *Nihongo-wa doo kawaru-ka: Goi to moji* [How does the Japanese language change? Words and characters]. Tokyo: Iwanami Shoten.

KAIHO, H. (1989). Joohoo-o hakaru: Entoropii, joohoo dentatsu-ryoo, joochoodo [Measuring information: Entropy, the amount of transmitted information, and redundancy]. In H. Kaiho (Ed.), *Shinri kyooiku đeeta-no kaiseki-hoo 10-koo: Ooyoo* [10 lectures on data analysis in psychology and education: Application] (pp. 14-26). Tokyo: Fukumura Shuppan.

KAMATA, T. (1991). *Kuwashii shoogakkoo kanji jiten* [A detailed dictionary of elementary school kanji]. Tokyo: Bun'eido.

KATO, M. (1989). Gakushuu kanji seigen-an oyobi gen "Jooyoo Kanji"-o meguru shojikoo ichiran-hyoo [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.

KESS, J. F., & MIYAMOTO, T. (1999). *The Japanese mental lexicon: Psycholinguistic studies of kana and kanji processing*. Philadelphia: Benjamins.

MINISTRY OF EDUCATION, CULTURE, SCIENCE, SPORTS, AND TECHNOLOGY, GOVERNMENT OF JAPAN (1987). *Shoogakkoo shidoosho: Kokugo-hen* [The Japanese language: The course of study at elementary school]. Osaka: Osaka Shoseki.

MINISTRY OF EDUCATION, CULTURE, SCIENCE, SPORTS, AND TECHNOLOGY, GOVERNMENT OF JAPAN (1998). *Monbushoo kokuji: Shoogakkoo gakushuu shidoo yooryoo* [The announcement of the elementary school course of study by the Ministry of Education, Culture, Science, Sports, and Technology, Government of Japan.]. Tokyo: Gyoosei.

NATIONAL INSTITUTE FOR JAPANESE LANGUAGE (1976). *Gendai Shinbun-no Kanji* [Japanese kanji characters in modern newspapers]. Tokyo: National Institute for Japanese Language.

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**, 44-55.

NOMURA, M. (1989). Kanji-no zoogo ryoku [A power of kanji productivity]. In K. Sato (Ed.), *Kanji kooza: Vol. 1. Kanji towa* [Kanji lecture series: Vol. 1. What is kanji?] (pp. 193-217). Tokyo: Meiji Shoin.

SHANNON, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379-423 (Pt. I) and 623-656 (Pt. II).

SNODGRASS, J. G., & MINTZER, M. (1993). Neighborhood effects in visual word recognition: Facilitatory or inhibitory? *Memory & Cognition*, **21**, 247-266.

TAKASHIMA, T. (2001). *Kanji to nihon'jin* [Kanji and Japanese]. Tokyo: Bungei Shunju.

TAMAOKA, K. (1991). Psycholinguistic nature of Japanese orthography. *Gengo Bunka Kenkyuu* [Studies in Language and Literature] (Matsuyama University), **11**, 49-82.

TAMAOKA, K. (2003). Where do statistically-derived indicators and human strategies meet when identifying On- and Kun-readings of Japanese kanji? *Cognitive Studies*, **10**, 1-28.

TAMAOKA, K., & ALTMANN, G. (2004). Symmetry of Japanese kanji lexical productivity on the left- and right-hand sides. *Glottometrics*, **7**, 65-84.

TAMAOKA, K., KIRSNER, K., YANASE, Y., MIYAOKA, Y., & KAWAKAMI, M. (2002). A Web-accessible database of characteristics of the 1,945 basic Japanese kanji. *Behavior Research Methods, Instruments, & Computers*, **34**, 260-275.

TAMAOKA, K., MIYAOKA, Y., & LIM, H. (2003). Entoropii to joochoodo de hyoogen-no tayoosei to kisokusei-o arawasu kokoromi: Kankokugo-kei nihongo gakushuusha no keigo hyoogen-o rei-ni [Measuring variation and regularity of expressions using entropy and redundancy: An example of politeness expressions used by Korean speakers learning Japanese]. *Nihongo Kagaku* [Japanese linguistics], **14**, 98-112.

VAN HEUVEN, W. J. B., DIJKSTRA, T., & GRAINGER, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory & Language*, **39**, 458-483.

YASUNAGA, M. (1981). Jooyoo kanji-hyoo ga umareru-made [A background history of the Jooyoo Kanji list]. *Gengo seikatsu* [Language life], **355**, 24-31.

YOKOYAMA, S., SASAHARA, H., NOZAKI, H., & LONG, E. (1998). *Shinbun denshi media-no kanji: Asahi shinbun CD-ROM-ni yoru kanji hindo hyoo* [Japanese kanji in the newspaper media: Kanji frequency index from the Asahi newspaper on CD-ROM]. Tokyo: Sanseido.

ARCHIVED MATERIALS

The database for 1,945 basic Japanese kanji, fourth edition, which was introduced in the article, may be accessed through the Psychonomic Society's Norms, Stimuli, and Data archive, http://www.psychonomics.org/archive/.

To access this file, search the archive for this article using the journal (*Behavior Research Methods, Instruments, & Computers*), the first author's name (Tamaoka), and the publication year (2004).

FILE: Tamaoka-BRMIC-2004b.zip.

DESCRIPTION: The compressed archive file contains one Excel file of the fourth edition kanji database, 1945kanji4th.xls.

AUTHOR'S E-MAIL ADDRESS: ktamaoka@hiroshima-u.ac.jp.

(Manuscript received January 5, 2004;
revision accepted for publication July 30, 2004.)