Cognitive Studies, **10**(4), 441-468. (Dec. 2003)

● Research Paper ●

Where Do Statistically-Derived Indicators and Human Strategies Meet When Identifying On- and Kun-Readings of Japanese Kanji?

Katsuo Tamaoka

The present study investigated attributes of kanji On- and Kun-readings from the perspectives of both statistical prediction and human strategy. In Study 1, discriminant analysis using the stepwise method revealed four significant indicators out of ten kanji characteristics for distinguishing On- and Kun-readings. These indicators are semantic concreteness, naming latency, special sounds and number of strokes. In Study 2, an On- or Kun-reading test is given to 30 native Japanese speakers. The result showed tendencies similar to the accuracy rates of discriminant analysis. After the test, a questionnaire revealed that 6 out of 10 strategies were employed by more than 6 out of the 30 participants. Three of these were congruent with significant indicators specified by discriminant analysis, namely, semantic concreteness, naming latency and special sounds. Despite the significant indicator in Study 1, particular strategies concerning kanji strokes and radical frequency were not used by humans. Native Japanese speakers are likely to use kanji neighborhood, kanji homophones and number of morae. The results between indicators and strategies illustrate a more general point: On- and Kun-readings can be effectively predicted by discriminant analysis on the basis of various kanji characteristics; however, due to a lack of consistency in On- and Kun-readings attached to each kanji, humans can flexibly incorporate a wider variety of strategies when making their determinations.

Keywords: Japanese kanji, On- and Kun-readings, discriminant analysis, human strategy

1. Introduction

Japanese *kanji* were adopted from Chinese characters to provide the Japanese with a way to write down their spoken language (see details about kanji in Hadamitzky & Spahn, 1981; Nomura, 1988; Kabashima, 1989; Kess & Miyamoto, 1999; Miller, 1967; Seeley, 1984a; Takamatsu, 1997; Takashima, 2001; Tamaoka, 1991; Tsukishima, 1979; Wydell, Butterworth & Patterson, 1995). Kanji pronunciation can be divided into two types; *On*-readings derived from the original Chinese pronunciation, and *Kun*-readings originating from the Japanese pronunciation. Native Japanese speakers are able to judge On-readings and Kun-readings at a relatively high probability rate when they listen to kanji pronunciations. It is difficult to imagine an arbitrary mapping between kanji and their On- and Kun-readings constructed in the lexicon of native Japanese speakers without rules.

This study investigated strategies used by native speakers to identify On- and Kun-readings. This question was approached from two perspectives. First, from a linguistic perspective, Study 1 employed discriminant analysis to identify and explore kanji characteristics based on the assumption that some kanji characteristics effi-

Institute for International Education, Hiroshima University.

ciently distinguish On- and Kun-readings. Second, from a human strategic perspective, Study 2 investigated the nature of considerations made by native Japanese speakers to distinguish between On- and Kun-reading. Approached from both perspectives, the present study further clarified similarities and differences between significant indicators and human strategies.

2. Phonological Sense of On- and Kun-readings Held by Native Japanese Speakers

There is evidence that native Japanese speakers form a phonological boundary between On-reading and Kun-reading. Tamaoka (2002) showed that kanji with a 50 ± 5 percent of On- and Kun-reading ratio were likely to be pronounced in On-reading when randomly embedded in kanji with a high On-reading ratio. In contrast, kanji were pronounced as Kun-readings when embedded in those with a high Kun-reading ratio. For example, the kanji 歌 meaning 'song' is pronounced in its On-reading /ka/ by 71.74 percent of 60 native Japanese speakers when embedded in kanji with a high On-reading ratio such as 郡 識 将 議 容 歌 銀 純 律 (the underlined kanji indicates the target kanji). Likewise, 歌 is pronounced as Kun-reading /uta/ at the rate of 97.83 percent when embedded in kanji with a high Kun-reading ratio such as 舌 顔 切 仲 割 歌 読 机 孫. Taking this as evidence, the study suggested that a kanji phonological lexicon must have On and Kun phonological boundaries as well as On and Kun sound tags.

Early studies on On- and Kun-readings by Nomura (1978, 1979) found that kanji with a high On-reading ratio (i.e., a low Kun-reading ratio) were named faster than those with a low On-reading ratio (i.e., a high Kun-reading ratio). The dual route model (e.g., Coltheart, 1985, 1987; Coltheart, Davelaar, Jonasson, & Besner, 1977; Coltheart, Curtis, Atkins & Haller, 1993; and later adapted by Goryo, 1987; Saito, 1981 for Japanese), which assumes two lexical routes of direct and indirect access via phonology to reach kanji semantics, was used to explain his results. Nomura (1978, 1979) and Kaiho and Nomura (1983) later interpreted that processing kanji with Kun-readings involves the activation of semantic representations before being named while kanji with On-readings were named by direct activation of phonological representations.

Based upon this series of studies, a strategy to distinguish On- and Kun-readings was provided as follows: When kanji meanings can be orthographically determined faster than sounds, the kanji is judged as a Kun-reading. On the contrary, when kanji sounds can be determined more easily and quickly than meanings, they are judged as On-readings. Nomura (1978, 1979), looking at speed of naming rather than On or Kun judgments, found that On-reading was activated more quickly than Kun-reading. His results suggested that On-reading can perhaps be distinguished from a Kun-reading simply because the former activates pronunciation quicker, showing that the degree of phonological accessibility acts as an important indicator in identifying On- and Kun-readings.

3. Use of On- and Kun-Readings for Identifying Kanji Sounds: The Linguistic Perspective

Usage of On- and Kun-readings relates to word origin. For example, the kanji 空 ('sky'), is pronounced as a single syllable or two morae /kuR/ as On-reading. On-reading is typically used for kanji compound words such as 空気 /kuR ki/ ('air'), 空中 /kuR tyuR/ ('in the air'), 空港 /kuR koR/ ('airport'), and 空想 /kuR soR/ ('fantasy'). A majority of kanji compound words with On-reading were created in the late Edo (1603–1867) and Meiji (1868–1912) Periods for the purpose of translating into Japanese new words from foreign books written in alphabetic scripts (Kabashima, 1989; Takashima, 2001). These words *are often* referred to as 近代新語/kiN dai siN go/ ('new modern words').

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

Number Kun-readings	Number of On-readings 0	1	2	3	4	5	Total	Percent
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%
Percent	2.06%	84.47%	12.54%	0.87%	0%	0.05%	100.00%	

Note 1: Data for this table was taken from the *Database for the 1,945 Basic Japanese Kanji* (Tamaoka, Kirsner, Yanase, Miyaoka & Kawakami, 2001, 2002).
Note 2: Rare pronunciations used for *wago* and nouns of proper places and names were excluded.

On the other hand, the kanji 空 /sora/, in Kun-reading, is a single word originating from the traditional Japanese vocabulary called *wago* 和語. A majority of kanji used as single, free-standing lexical units originated from *wago* and so are read in Kun-reading. Thus, native Japanese speakers likely use Kun-reading for a single free-standing noun presented in a single kanji. *Wago* is also written using two kanji such as 空耳 /sora mimi/ ('mishearing') and 空言 /soragoto/ ('falsehood'). As such, On- and Kun-readings are used distinctly for different words: On-readings for *kango* 漢語 (traditional Chinese words) and Kun-readings for *wago* (traditional Japanese words).

Tanaka (1978) introduced a study on the frequency of the printed occurrence of 38,395 different words as calculated by the National Institute for Japanese Language. *Kango* occupied 47.5 percent of the Japanese vocabulary whereas *wago* comprised 36.7 percent. *Gairaigo*, derived from alphabetic languages, made up only 9.8 percent and *konshugo* (a mixture of different types), only 6.0 percent. As such, more than 85 percent of the total Japanese vocabulary consisted of *kango* and *wago*. Usage of On- and Kun-readings of kanji fundamentally differed depending upon their etymology; in other words, whether they were considered *kango* or *wago*. It is useful to have sense of On- and Kun-readings to identify sounds of kanji compound words; otherwise, pronunciations have to be memorized for each word.

On- and Kun-reading distinction, however, is used not only for free-standing nouns, but also verbs, adjectives and adverbs. Kanji describe a semantic part of these words (Sato, 1996). For example, the verb 歩く /aru(ku)/ is constructed by kanji and hiragana. The kanji 歩 /aru/ indicates the semantic concept 'walk' while the hiragana く /ku/ shows a verb inflection. Likewise, the adjective 美しい /utuku(sii)/ ('beautiful') consists of kanji and hiragana. The kanji 美 /utuku/ refers to the concept 'beauty' while the hiragana しい /sii/ refers to an adjective inflection. Almost all these verbs, adjectives and adverbs are pronounced in Kun-reading with a combination of hiragana sounds. Kun-reading for these words seems to be automatically selected by native Japanese speakers. Again, it is troublesome to find proper sounds without a clear sense of On- and Kun-readings.

4. Ratios of On- and Kun-readings in Basic Japanese Kanji

In order to establish a standard for kanji usage, in 1981 the Japanese government published the *Joyo Kanjihyo*, 'The List of 1,945 Commonly-used Basic Kanji', officially simplifying the number of kanji and their pronunciations in (see details in Kato, 1989; Ministry of Education, Culture, Science, Sports and Technology, Government of Japan, 1978, 1987, 1998; Seeley, 1984b; Tamaoka, 1991; Watanabe, 1989). Using this list, Tamaoka, Kirsner, Yanase, Miyaoka and Kawakami (2001) produced the *Database for the 1,945 Basic Japanese Kanji* (2nd edition), in which 30 columns of differing kanji characteristics such as kanji print frequency, kanji homophones, On-reading ratio and neighborhood size are recorded. While it is generally believed that all Japanese kanji have both On- and Kun-readings, only 1,169 kanji (60.10%) of the commonly-used 1,945 actually have both pronunciations. As shown in Table 1, the Database shows that 32 kanji out of 1,946 have a single On-reading and no Kun-reading, whereas 667 kanji have a single Kun-reading and no On-reading. As for kanji with multiple On- and Kun-readings, there are 40 with only Kun-readings (2.06%) and 737 with only On-readings (37.84%).

The number of kanji with only On-readings exceeded that with only Kun-readings by more than 18 times. This great difference is to be expected since the reason for such a limited number of kanji with a single Kun-reading is due to the fact that Kun-readings were created in Japan, while kanji borrowed from Chinese kept a lot of their On-readings, kanji assigned to traditional Japanese words were given Kun-reading and lost their On-reading.

5. Possible Kanji Characteristics Identifying On- and Kun-readings

In order to identify kanji characteristics which distinguish On- and Kun-readings, the present study selected ten possibilities as outlined below.

According to studies by 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, while the second variable is the naming latency of kanji. Studies done by Nomura (1978, 1979) and Hirose (1998) suggest that kanji pronounced using Kun-readings are likely to be more independent and concrete in their meaning. Nomura (1978, 1979) calculated On-reading ratios based on the frequency of printed kanji provided by the National Institute for Japanese Language (Kokuritsu Kokugo Kenkyujo, 1962) and found that kanji with a low On-reading ratio (i.e., a high Kun-reading ratio) were pronounced slower than those with a high On-reading ratio (i.e., a low Kun-reading ratio), explaining that kanji with Kun-readings were related to their meaning causing an activation of semantic representations before the activation of phonological representations, even though the task required only vocalization of kanji pronunciation. There are some exceptional cases to this semantic attachment, as Nomura (1978) himself noted in his paper. For example, the kanji 肉 refers directly to the concrete concept of 'meat', but its pronunciation of /niku/ is an On-reading. Many native Japanese speakers confuse such exceptional kanji as having Kun-readings instead of On-readings (Kayamoto, 2000). Despite this, Kun-readings are generally much more closely attached to meaning than On-readings (Kaiho & Nomura, 1983).

Distinctions between On- and Kun-readings for kanji can be supported from a phonological as well as semantic viewpoint. In dispensing with Chinese tones, Japanese created multiple readings for kanji. For example, the sound /ka/ can be written by various kanji such as 火 (fire), 化 (chemical), 科 (section), 花 (flower), 家 (house) and 蚊 (mosquito). All these readings are exclusively On-readings. According to the *Database for the 1,945 Basic Japanese Kanji* (2nd edition) created by Tamaoka et al. (2001), the most shared sound of all kanji was the On-reading of /syoR/, found in 65 kanji. This is approximately 3.3 percent of the total number of basic kanji. The sound of /koR/, having 65 On-readings and one Kun-reading, also showed the same number of homophones as /syoR/. The sounds of almost all these homophones derived from On-readings.

With such a great number of kanji sharing the same sound in Japanese, On-readings cannot in most cases be attached to a single meaning. The characteristics of multiple kanji with On-readings could then be a third possible variable to distinguish between On- and Kun-readings.

A fourth possible variable is 'kanji neighborhood size', which represents the number of times one unit of kanji can combine with another to create two-kanji compound words. For example, 学 /gaku/, ('learning' or 'to learn'), can be found in various kanji combinations such as 学術 /gaku zyutu/ (academic), 学生 /gaku sei/ (student), 学歴 /gaku reki/ (one's educational background), 大学 /dai gaku/ (university) and 数学 /suR gaku/ (mathematics). All these two-kanji compound words are pronounced using On-readings, so it is expected that if On-readings are often used for compound words produced by two or more kanji, those with a larger neighborhood size will likely have a higher number of On-readings. In other words, kanji with a larger neighborhood size may be pronounced more frequently using On-readings than Kun-readings.

A fifth possible variable for distinguishing between On- and Kun-readings is whether or not a kanji contains the special sounds of /N/, /R/ and /Q/. Despite its large scale adoption of Chinese characters, Japanese is quite different in terms of its sound system. The Japanese sound system does not have the four tones which exist in Mandarin, the standard form of the Chinese language. Furthermore, Japanese in the Nara period (the years from 710 to 784) 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 CVC syllabic structure such as /teN/, /keN/, /haQ/, /kaQ/. These syllables began to appear in the Japanese sound system from the Heian period (794 A.D. to 1192 A.D.) when kanji was adopted from Chinese (Komatsu, 1981; Kubozono & Ota, 1998; Numoto, 1987). Kanji phonological characteristics of special sounds could be a possible characteristic used by native Japanese speakers to identify On- and Kun-readings.

Table 2 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
	None	851
2	/R/ and /Q/	3
	/R/ and /N/	1
	/R/	91
	/N/	40
	None	109
3	/R/ and /Q/	2
	/R/ and /N/	1
	/R/	6
	/N/	2
	None	6
4	-	0
5	/R/, /N/ and /Q/	1
	Total	1,945
Kanji with special sounds		939
Kanji with no special sounds		966
Ratio of special sounds		49.29%

Note: Data for this table was taken from the *Database for the 1,945 Basic Japanese Kanji* (Tamaoka, Kirsner, Yanase, Miyaoka & Kawakami, 2001).

As shown in Table 2, the *Database for the 1,945 Basic Japanese Kanji* shows that 939 out of 1,905 kanji have special sounds in their On-readings. (It should be noted that 40 of the 1,945 basic kanji do not have On-readings. Thus, only 1,905 kanji were taken into account.) Since 966 out of 1,905 kanji do no have these special sounds, the ratio of special sounds inclusive in On-readings is calculated as 49.29 percent. In contrast, as shown in Table 3, 48 out of 1,209 kanji with Kun-readings contain special sounds (9 kanji have two Kun-readings, while 736 do not have Kun-readings.) A majority of kanji with Kun-readings (96.06 percent of 1,170 kanji) do not have special sounds in their pronunciations. Historically speaking, since special sounds were introduced to the Japanese sound system following the Heian period, it is natural to find many special sounds in *kango* which use On-readings.

The five remaining candidates to be considered as possible variables in distinguishing On- and Kun-readings are: kanji frequency in Japanese printed materials, number of morae in the length of phonological structure, kanji radical frequency of the 1,945 basic kanji, number of strokes required to write a kanji, and school grade indicating age of kanji acquisition.

Two of the ten candidates, naming latency and semantic concreteness, were measured using native Japanese speakers. Data for naming latency was taken in an experimental setting while data for semantic concreteness was obtained by questionnaire (details are explained in *stimuli* in Experiment 1). The remaining eight kanji characteristics were taken from the index of the kanji database (Tamaoka et al., 2001). Among these, special sounds, number of morae and number of kanji homophones are phonological characteristics of kanji while radical frequency and number of strokes are orthographic characteristics. These multiple kanji characteristics are taken into consideration in order to determine possible candidates to identify On- and Kun-readings.

6. Outline of Study

The present study approached the question of On- and Kun-reading selection from two perspectives. Study 1 employed a statistical technique called 'discriminant analysis' to distinguish possible indicators of On- and Kun-readings based on ten variables related to kanji (see Appendix). The research question addressed in Study 1 focused on the degree to which each variable could differentiate On-reading from Kun-reading. However, whether or not native Japanese speakers in fact use these particular kanji characteristics to identify On- and Kun-readings is a different issue. Therefore, those kanji items not correctly classified by discriminant analysis in Study 1 were further tested using native Japanese speakers in Study 2. These incorrectly classified kanji were randomly listed with correctly classified kanji and presented to native Japanese speakers who were asked to identify On- or Kun-readings. In addition, after the On- and Kun-reading test, the native Japanese speakers were asked which characteristics they used to determine On- and Kun-readings. In Study 2, the focus was whether or not native Japanese speakers displayed the same tendencies as indicated by the discriminant analysis of the first study. Finally, the results of Study 1 and Study 2 were compared to clarify and identify mechanisms of On- and Kun-reading used by native Japanese speakers.

Table 3 Special Sounds /N/, /R/, /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
	None	743
2	/R/	4
	/N/	2
	None	288
3	/R/	4
	/N/	1
	/Q/	1
	/R/ and /N/	1
	None	87
4	/Q/	1
	/N/ and /R/	3
	/N/ and /Q/	1
	None	37
5	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
Ratio of special sounds		3.94%

Note 1 : Data for this table was taken from the *Database for the 1,945 Basic Japanese Kanji* (Tamaoka, Kirsner, Yanase, Miyaoka & Kawakami, 2001).

Note 2 : Nine kanji had special sounds twice in their Kun-readings. Thus, the total number of spcial sounds was 1,954 for the 1,945 basic kanji..

7. Study 1 — Predicting On- and Kun-readings of Kanji from Ten Characteristics

7.1 Method

Selection of Kanji. Using the *Database for the 1,945 Basic Japanese Kanji*, a stratified sample of 21 out of 32 kanji with a single Kun-reading and 113 out of 667 kanji with a single On-reading were selected for discriminant analysis (see Table 1). All 134 actual kanji items used in this study are listed in the Appendix.

Ten Possible Indicators of Kanji Reading. The following ten characteristics were used in the discriminant analysis. (All statistical data with regards to these ten characteristics are listed in the Appendix.)

(1) Degree of Semantic Concreteness:

Forty-eight Japanese undergraduate students, (different from the group which participated in measuring naming latency) participated in rating semantic concreteness of 134 kanji. The overall average age of participants was 21 years and 2 months with a standard deviation of 1 year and 3 months. Degree of concreteness for each kanji's meaning was measured on a seven-point scale, from 1 ('very abstract') to 7 ('very concrete'). An average of ratings for each kanji was recorded for analysis. Extreme discrepancies in ratings were seldom seen throughout the 134 kanji. The kanji 愛 ('love') was rated a mean of 4.88 with the largest standard deviation of 2.15. No other kanji showed a standard deviation larger than 2.00.

(2) Naming Latency:

Twenty-three Japanese undergraduate students, (different from the group which rated semantic concreteness) participated in a naming task of 134 kanji. The overall average age of participants was 20 years and 7 months with a standard deviation of 1 year and 1 month. Individual stimulus items were randomly presented to participants in the center of a computer screen (Toshiba, J-3100 Plasma display), 600 milliseconds after the appearance of an asterisk '*' indicating an eye fixation point. Participants were instructed to pronounce items as quickly and as accurately as possible. The present study selected only kanji with a single pronunciation, controlling for multiple readings of kanji so as not to affect naming latency (Kayamoto, Yamada & Takashima, 1998). The means of naming latencies for the kanji are reported in the Appendix. Accuracy ratios of the 134 kanji highly correlated with those of naming latencies [$n = 134$, $r = -0.64$, $p < .0001$]: the lesser the error rates, the faster the naming latencies. Since a multi-correlation between naming latencies and accuracy ratios was observed in the discriminant analysis, the variable of accuracy ratios was excluded from being considered a possible indicator of kanji reading.

(3) Neighborhood Size:

Kawakami (1997) provides an index of kanji neighborhood size which is also recorded in the *Database for the 1,945 Basic Japanese Kanji.* The term 'kanji neighborhood size' refers to the number of times one unit of kanji can combine with another kanji to create two-kanji compound words, which are produced by the combination of kanji placed in the left-hand and right-hand positions of a word. The present study considered both left-hand and right-hand positions in its calculation of kanji neighborhood size.

(4) Special Sounds:

To conduct the discriminant analysis, kanji with a special sound (i.e., /N/, /R/ and /Q/) was recorded '1' and kanji with no such sounds in either On- or Kun-reading '0'. Among the 134 kanji selected for the present study, there were none with the special sound of /Q/. There was only one kanji with a Kun-reading of 峠 /toRge/, ('a mountain pass') which has the special sound of a long vowel /R/. The kanji 峠 is actually one of the few kanji created by Japanese which uses the combination of 山 /yama/, ('mountain'), 上 /ue/, ('up') and 下 /sita/, ('down') (Kobayashi, 1998).

(5) Number of Kanji Homophones:

Pronunciation is often shared by multiple kanji. For example, the sound /yoku/ can be written using five different kanji out of the basic 1,945. Each of these five kanji with the same pronunciation are indicated by the number '5'. Both On- and Kun-readings were calculated for kanji homophones. The index of kanji homophones was taken from the *Database for the 1,945 Basic Japanese Kanji.*

(6) Kanji Frequency:

Yokoyama, Sasahara, Nozaki and Long (1998) published frequency of occurrence data based on all kanji appearing in the Tokyo edition of the *Asahi Newspaper* printed throughout 1993. Their index was used in the present study under kanji frequency.

(7) Number of Morae:

As a phonological reference, the number of morae constructing each kanji was recorded. For example, the kanji 列, ('line'), with the On-reading of /retu/ consists of two morae (/re/ and /tu/), so the number of morae was recorded as '2'. Likewise, the kanji 娘, ('daughter'), with a Kun-reading of /musume/ has three morae (/mu/, /su/ and /me/) so the number of morae was recorded as '3'.

(8) Radical Frequency:

Radical frequency indicates how many of the 1,945 basic kanji share the same radical. A large body of kanji (1,057 characters or 54.34% of the 1,945 basic kanji) is constructed using only 24 of a possible 214 radicals. Again, the *Database for the 1,945 Basic Japanese Kanji* provided the index of radical frequency for the 134 kanji used in this study.

(9) Number of Strokes:

The number of strokes required to write each kanji were taken from a Japanese kanji dictionary (Kamata, 1991) which is recorded in the *Database for the 1,945 Basic Japanese Kanji.*

(10) School Grade:

Figures shown for 1,006 kanji follow the Japanese language curriculum as established by the Ministry of Education, Culture, Science, Sports and Technology, Government of Japan in 1989. Since the remaining 939 kanji are taught in Grades 7-9 with no grades specifications, they were indicated with the number '7'. These school grade figures are recorded in the *Database for the 1,945 Basic Japanese Kanji.* The present study used school grade as an index of the age when these kanji are acquired.

7.2 Results

7.2.1 Correlation Analysis

Means, standard deviations and Pearson's correlation coefficients for the ten characteristics of the selected 134 kanji are shown in Table 4. There were no extremely high correlations among variables, suggesting that each measure of kanji remained relatively independent. Accuracy ratios of stimuli for naming kanji were excluded from the discriminant analysis due to its high correlation with naming latency.

As shown in Table 4, several variables showed relatively higher correlations. Variables which correlated significantly high at the level of 1 percent ($p < .01$) are discussed as follows:

Neighborhood size had a significant correlation with naming latency [$r = -0.354$, $p < .01$], kanji frequency [$r = 0.334$, $p < .01$] and school grade [$r = -0.500$, $p < .01$]. Kanji which can produce many two-kanji compound words (i.e., those with a larger neighborhood size) are named faster, likely to be used more frequently in print, and tend to be taught earlier in school.

Table 4 Means, Standard Deviations and Correlations for the Ten Characteristics of Kanji

Variables	1	2	3	4	5	6	7	8	9	10
1 Neighborhood size	-									
2 Special sounds (/N/ or /R/)	-0.013	-								
3 Naming latency (ms)	-0.354 **	0.029	-							
4 Number of morae	-0.030	0.258 **	0.175 *	-						
5 Number of kanji homophones	-0.065	0.347 **	0.011	-0.240 **	-					
6 Radical frequency	-0.051	-0.120	0.053	-0.128	0.045	-				
7 Number of strokes	-0.027	0.043	0.170 *	-0.021	0.093	0.149	-			
8 Kanji frequency	0.334 **	-0.049	-0.280 **	0.006	-0.033	-0.048	0.027	-		
9 Semantic concreteness	0.087	-0.104	-0.226 *	0.237 **	-0.376 **	0.016	0.031	-0.006	-	
10 School grade	-0.500 **	0.066	0.515 **	0.168	-0.084	0.025	0.135	-0.502 **	-0.078	-
Means	51.87	0.40	628	1.90	12.96	27.04	10.28	4828	4.79	5.32
Standard Deviations	37.01	0.49	86	0.37	14.61	25.40	3.31	7138	0.91	1.86

Note: n=134. * $p<.05$. ** $p<.01$.

The special sounds of /N/ and /R/ significantly correlated with number of morae [$r=0.212$, $p<.01$] and number of kanji homophones [$r=0.361$, $p<.01$]. Kanji with these special sounds are likely to have many morae and to share many homophones.

Kanji with shorter naming latencies tended to have a higher kanji frequency [$r=-0.280$, $p<.01$] and were taught at lower school grades [$r=0.515$, $p<.01$]. Naming latency refers to the length of time from the onset of kanji visual presentation to the beginning of vocalization. In other words, it indicates how quickly kanji phonology is activated. When kanji are frequently used in print and taught at lower school grades, they are likely to be seen more often in daily life. As a result, these kanji tend to be named faster.

The number of morae showed a significantly negative correlation with the number of kanji homophones [$r=-0.240$, $p<.01$] and the degree of semantic concreteness [$r=0.237$, $p<.01$]. The longer the moraic string of kanji, the lesser the homophones. Since kanji with a longer moraic structure can have wider variations, it is natural to have a negative correlation with the variable of kanji homophones. A longer moraic structure showed a positive correlation with the degree of semantic concreteness.

The number of kanji homophones showed a significant negative correlation with the degree of semantic concreteness [$r=-0.376$, $p<.01$]. The larger the number of kanji homophones, the lesser the semantic concreteness. This result is interesting as it shows that kanji with many homophones create semantic ambiguity.

Finally, kanji frequency showed a significant correlation with school grade [$r=-0.502$, $p<.01$]. Stimuli with a high kanji frequency are likely to be taught in the lower grades, suggesting that kanji taught at the lower grades may be used more frequently in printed materials.

7.2.2 Discriminant Analysis and Post-hoc Analysis

Discriminant analysis using all ten kanji variables was conducted on all 134 items to find out if these could serve to sufficiently distinguish between On- and Kun-readings. Since scales of the ten variables (or kanji characteristics) greatly differ, a normalized score (z-score) was used to control all variables as the mean of 0 with 1 standard deviation. Post-hoc analyses were conducted using the original scores.

As shown in Table 5, the analysis correctly classified kanji readings for 116 of 134 kanji (86.57 percent of the kanji in total). Out of 113 kanji, 97 (85.84%) were correctly classified as having On-readings. Similarly, 19 out 21 kanji (90.48%) were correctly classified as having Kun-readings. Consequently, these ten kanji characteristics were able to distinguish On-readings from Kun-readings for the 134 selected kanji with a high success rate of over 85 percent. The variables of this analysis are listed in order of how accurately they could predict kanji reading (see Table 6).

As shown in Table 6, there were four signifi-

Table 5 Classification Results for On- and Kun-readings

Actual Group	Predicted Group: Kun-readings	Predicted Group: On-readings	Total
Kun-readings	19	2	21
(n=21)	90.48%	9.52%	100%
On-readings	16	97	113
(n=113)	14.16%	85.84%	100%

Note 1 : Correctly classified cases were 116 out of 134 cases or 86.57 percent.
Note 2 : Shaded columns indicate correctly classified kanji.

Table 6 Results of Discriminant Analysis for Predicting On- and Kun-readings of Kanji

#	Variables	Discriminant Function Coefficients	Mahalanobis' Distance	F-value
1	Semantic concreteness	1.7149	3.1786	21.5495 ****
2	Naming latency	1.2537	4.1000	10.0216 **
3	Radical frequency	0.4323	4.8664	1.7470
4	School grade	0.3746	4.9714	0.6925
5	Number of morae	0.3047	4.9741	0.6655
6	Kanji frequency	-0.1972	5.0139	0.2711
7	Neighborhood size	-0.2710	4.9888	0.5192
8	Number of kanji homophones	-0.4489	4.9083	1.3241
9	Number of strokes	-0.9056	4.3397	7.3174 *
10	Special sounds (/N/ or /R/)	-1.1253	4.1680	9.2433 **

Note1 : n=134. * $p<.05$. ** $p<.01$. *** $p<.001$. **** $p<.0001$.
Note2 : The above 10 variables are converted to z-scores for discriminant analysis.

cant indicators of On- and Kun-readings: degree of semantic concreteness, naming latency, special sounds and number of strokes. The other six variables of radical frequency, number of kanji homophones, school grade, number of morae, neighborhood size and kanji frequency were not significant indicators of On- and Kun-readings. The results of the ten variables relation to kanji characteristics are discussed below.

(1) Degree of Semantic Concreteness

The most significant indicator was the degree of semantic concreteness [$F=21.5495$, $p<.0001$]. In fact, there was a large difference in the means of semantic concreteness for On-readings (n = 113, M = 4.64, SD = 0.86) and Kun-readings (n = 21, M = 5.58, SD = 0.79) upon conducting a post-hoc analysis with a t-test [n = 134, $t(132)=4.670$, $p<.0001$]. Kanji with Kun-readings proved to be more concrete in their meanings than kanji with On-readings, as suggested by the previous studies of Nomura (1978, 1979) and Kaiho and Nomura (1983).

(2) Naming Latency

Naming latency was the second most significant indicator [$F=10.0216$, $p<.01$]. These results were consistent with the experimental results of Nomura's studies (1978, 1979), which found that kanji with high Kun-reading ratios (i.e., a low On-reading ratio) take longer to name than those with low Kun-reading ratios (i.e., a high On-reading ratio). The explanation given by Nomura (1978, 1979) and Kaiho and Nomura (1983) is that kanji with Kun-readings are strongly attached to semantics while kanji with On-readings are more attached to phonology. The t-test also showed a significant difference in the means of naming latencies between kanji with On-readings (n = 113, M = 620ms, SD = 87ms) and those with Kun-readings (n = 21, M = 676ms, SD = 63ms) [n = 134, $t(132)=2.844$, $p<.005$].

(3) Special Sounds

As was anticipated, special sounds were the third most significant indicator [$F=9.2433$, $p<.01$]. In fact, there was only one kanji with a single Kun-reading containing the special sound of /R/ (i.e., 峠 /toRge). In contrast, 52 out of 113 kanji with On-readings (53.98%) contained the special sounds of /N/ or /R/; more precisely, there were 26 kanji with /N/ (23.01%) and 26 with /R/ (23.01%). Therefore, this result suggests that kanji pronunciation including special sounds, especially /N/ and /R/, can be classified as On-readings while those with no special sounds can be identified as Kun-readings.

As discussed in the introduction of this study, after adopting Chinese characters the Japanese modified those having tones and those with final consonants of a CVC (i.e., ending with /n/ and /ng/) and possibly CVV phonological structure. This simplification process resulted in the creation of certain Japanese kanji with many special sounds and kanji homophones, as was shown by the t-test in the post-hoc analysis. Among kanji with On-readings ($n=113$), those with special sounds ($n=52$, $M=19.56$, $SD=16.88$) had a significantly greater number of kanji homophones than those without special sounds ($n=61$, $M=11.46$, $SD=11.87$) [$t(111)=2.98$, $p<.005$]. Consequently, kanji with special sounds are likely to have many homophones.

(4) Number of Strokes

The fourth most significant indicator of kanji reading was number of strokes [$F=7.317$, $p<.01$]. This result was unexpected since it had been assumed that visual complexity would have no bearing on distinguishing between On- and Kun-readings. Indeed, it is interesting to note that according to the t-test, the means of the number of strokes required to write kanji with On-readings ($n=113$, $M=10.47$, $SD=3.20$) versus those with Kun-readings ($n=21$, $M=9.23$, $SD=3.74$) did not differ significantly [$n=134$, $t(132)=-1.574$, n.s.].

(5) Radical Frequency

Radical frequency was not a significant indicator of On- and Kun-readings, and therefore did not seem to be related to any phonological aspects of kanji reading. The post-hoc analysis performed with the use of a t-test also did not show significant differences in the means of radical frequency between kanji with On-readings ($n=113$, $M=25.85$, $SD=24.67$) and kanji with Kun-readings ($n=21$, $M=33.48$, $SD=28.78$) [$n=134$, $t(132)=1.267$, n.s.].

(6) Kanji Homophony

Contrary to what had been expected, the number of homophones was not a significant indicator of kanji reading. As seen in the Appendix, kanji with Kun-readings did not have any homophones. In contrast, the means of kanji homophones with On-readings ($n=113$, $M=15.19$, $SD=14.89$) was much higher than for those with Kun-readings ($n=21$, $M=1.00$, $SD=0.00$). The t-test showed a significant difference in the means of the number of kanji homophones [$n=134$, $t(132)=-4.354$, $p<.0001$]. Due to a high standard deviation of kanji homophones with On-readings, it is expected that some kanji with On-readings have few homophones such as characters with Kun-readings. Therefore, although kanji homophones are frequently seen among others with On-readings, this variable does not seem to be a reliable indicator of On- and Kun-readings in light of the results of the discriminant analysis.

(7) School Grade

School grade was added as an index of the age at which certain kanji are acquired. Again, although a t-test revealed that kanji with On-readings ($n=113$, $M=5.14$, $SD=1.84$) were taught at significantly lower grades than those with Kun-readings ($n=21$, $M=6.29$, $SD=1.71$) [$n=134$, $t(132)=2.649$, $p<.01$], school grade was not a significant indicator for distinguishing between On- and Kun-readings.

(8) Number of Morae

The means of the number of morae for kanji with On-readings ($n = 113$, $M = 1.88$, $SD = 0.33$ and with Kun-readings ($n=21$, $M=2.05$, $SD=0.59$) were significantly different [$n=134$, $t(132)=1.998$, $p < .05$]. While Kanji with Kun-readings are likely to have a longer moraic structure than those with On-readings, this variable was not a significant indicator of On- and Kun-readings.

(9) Neighborhood Size

As discussed in the introduction, it had been expected that neighborhood size would be a strong candidate as a classifier of On- and Kun-readings. However, the discriminant analysis showed that neighborhood size was not significant. By way of t-test, the post-hoc analysis indicated that the means of neighborhood size between kanji with On-readings ($n = 113$, $M = 54.59$, $SD = 38.42$) and kanji with Kun-readings ($n = 21$, $M = 37.19$, $SD = 24.01$) were significantly different [$n = 134$, $t(132) = -2.001$, $p < .05$]. Thus, neighborhood size is larger among kanji with On-readings than kanji with Kun-readings, but this characteristic does not contribute to distinguishing between On- and Kun-readings.

(10) Kanji Frequency

The variable of kanji frequency is known to have a strong influence on the cognitive processing of compound words measured by various tasks (e.g., Balota & Chumbley, 1984; Forster, 1979; Grainger, 1990; Monsell, Doyle, & Haggard, 1989; Taft, 1979, 1991; Tamaoka & Hatsuzuka, 1995; Whaley, 1978). The discriminant analysis indicated that kanji frequency was the weakest indicator of On- and Kun-readings among the ten kanji characteristics, even though a t-test showed a significant difference in the means of kanji frequency between kanji with On-readings ($n = 113$, $M = 5373.09$, $SD = 7559.42$) and those with Kun-readings ($n = 21$, $M = 1894.48$, $SD = 2841.77$) [$n = 134$, $t(132) = -2.076$, $p < .05$].

7.3 Discussion of Study 1

Study 1 examined ten variables of kanji which could serve as significant indicators of On- and Kun-readings. Discriminant analysis found four significant indicators: degree of semantic concreteness, naming latency, special sounds and number of strokes. In light of the findings of previous studies (Kaiho & Nomura, 1983; Nomura, 1978, 1979), it was expected that semantic concreteness and naming latency would be significant indicators for On- and Kun-readings. The remaining six of the ten characteristics did not contribute in distinguishing On- from Kun-readings; namely, radical frequency, number of kanji homophones, school grade, number of morae, neighborhood size and kanji frequency. As such, via discriminant analysis, Study 1 provided a general list of kanji characteristics which identify On- and Kun-readings.

8. Study 2 — Identifying of On- and Kun-readings by Native Japanese Speakers

The discriminant analysis in Study 1 correctly classified 116 of 134 kanji (86.57%) into On- and Kun-readings. However, 18 out of 134 kanji items (16 kanji with On-readings and 2 with Kun-readings) were incorrectly classified. This led to the question of whether or not native Japanese speakers also have difficulty in properly classifying these 18 kanji into On- and Kun-readings. Therefore, Study 2 looked at how many readings native Japanese speakers could properly identify these 18 kanji. Subsequent to this, 10 questions related to the 10 kanji characteristics were asked to ascertain whether native Japanese speakers had used these characteristics as indicators for On- and Kun-readings during the test.

8.1 Methods

Participants. Thirty undergraduate students (28 males and 2 females), all native Japanese speakers participated in this experiment. Ages ranged from 18 years and 9 months to 20 years

Table 7 Means and Standard Deviations of On- and Kun-reading Decision for Three Different Types of Kanji

Category	Number of Kanji	*M*	*SD*
On-reading kanji	24	70.00%	14.43%
Kun-reading kanji	19	79.30%	12.55%
Incorrectly classified kanji	18	45.93%	23.66%

Note 1 : n=30. *M* = means. *SD* = standard deviations.
Note 2 : The incorrectly classified kanji consisted of 2 kanji with Kun-readings and 16 kanji with Kun-readings.

and 8 months, the average age being 19 years and 5 months with a standard deviation of 6 months on the day of testing.

Stimulus Items. Three different types of kanji stimuli were used in Study 2. Kanji were selected from those which could not be properly classified by the discriminant analysis. Eighteen kanji were chosen, 16 with On-readings and 2 with Kun-readings, forming the first type of stimuli. Kanji of the second type were those with Kun-readings which had been correctly identified by discriminant analysis, excluding two kanji whose Kun-readings had been misclassified as On- readings (i.e., 瀬 /se/ and 箱 /hako/). There were 19 kanji of this type. Kanji of the third type were selected from the list of kanji with On-readings (see Appendix). Using the stratified sampling method, every fourth kanji from a phonologically-ordered list which had been correctly identified by discriminant analysis was chosen. As a result, 24 kanji with On-readings were selected. The total number of kanji items was thus 61 (i.e., 18 incorrectly classified kanji items, plus 19 with Kun-readings and 24 with On-readings which had been correctly identified in Study 1).

Procedure. As can be seen below, the 61 items were randomly arranged on a single-sided, one page questionnaire such that each kanji, such as 滝 ('waterfall'), would have its pronunciation written in hiragana たき (/taki/) next to it. Participants were asked to place a check mark (√) in one of two boxes indicating On-reading (音読み) and Kun-reading (訓読み). In this manner, accuracy in On- and Kun-reading decisions for the three types of kanji (i.e., 24 with On-reading, 19 with Kun-reading and 18 incorrectly classified kanji) could be compared:

週 しゅう (/suR/)
☐ 音読み (On-reading)
☐ 訓読み (Kun-reading)
娘 むすめ (/musume/)
☐ 音読み (On-reading)
☐ 訓読み (Kun-reading)

Questions. After administering the test to identify On- and Kun-readings, 10 questions related to the use or non-use of 10 kanji characteristics for identifying On- and Kun-readings were posed to all 30 participants, who could check as many question items as they felt appropriate. These 10 questions are listed in Table 9.

8.2 Results

8.2.1 Test Identifying On- and Kun-readings

The means and standard deviations for the three different types of kanji are reported in Table 7. As expected from the results of Study 1, the 18 kanji whose readings had been incorrectly classified by discriminant analysis showed a mean of 45.93 percent for On- or Kun-reading decisions by native Japanese speakers. Because a correct answer was one of two choices (i.e., On-reading or Kun-reading), 50.00 percent was considered the level of random chance. In other words, even one who has no knowledge of kanji On- and Kun-readings could chose one of the two answers and by random chance still be *expected* to be correct 50.00 percent of the time. There-

Table 8 Eighteen Incorrectly Classified Kanji Items by Discriminant Analysis and Accuracy Rates of Native Japanese Speakers for On- and Kun-reading Decisions

#	Incorrectly Classified Items	Kanji Pronunciation	Kanji Meaning in English	Kanji Reading	On- or Kun-Decision (%)	Semantic Concreteness	Naming Latency (ms)	Special Sounds	Number of Strokes
1	肉	/niku/	meat	On	16.67%	5.88	549	-	6
2	服	/huku/	clothes	On	23.33%	5.92	599	-	8
3	枚	/mai/	a piece	On	23.33%	4.81	641	-	8
4	菊	/kiku/	chrysanthemum	On	26.67%	6.29	623	-	11
5	肺	/hai/	lung	On	26.67%	6.04	548	-	9
6	液	/eki/	liquid	On	30.00%	5.33	619	-	11
7	胃	/i/	stomach	On	36.67%	6.23	557	-	9
8	軸	/ziku/	axis	On	36.67%	5.10	649	-	12
9	陸	/riku/	land	On	36.67%	6.00	544	-	11
10	脈	/myaku/	pulse	On	40.00%	5.67	571	-	10
11	劇	/geki/	drama	On	46.67%	5.98	611	-	15
12	菌	/kiN/	germ	On	56.67%	5.56	751	/N/	11
13	尿	/nyoR/	urine	On	56.67%	5.77	615	/R/	7
14	酪	/raku/	farming	On	70.00%	3.79	904	-	13
15	才	/sai/	ability	On	86.67%	3.88	572	-	3
16	渓	/kei/	vallay	On	93.33%	3.67	1011	-	11
17	瀬	/se/	rapids	Kun	30.00%	4.10	701	-	19
18	箱	/hako/	box	Kun	90.00%	5.92	589	-	15

Note: n=30.

fore, a mean of 45.93 percent for incorrectly classifying the readings of the 18 kanji indicated that even native Japanese speakers could not properly identify these kanji. Since the participants were undergraduate students at a Japanese university, the mean of correct responses is considered to represent an approximately correct ratio of mature, native Japanese speakers of relatively-high academic accomplishment. In contrast, the participants were able to properly judge at 70.00 percent of the 24 kanji with On-readings and at 79.30 percent of the 19 kanji with Kun-readings. Accordingly, the means of correct ratio for On- and Kun-readings indicated a similar trend as shown by the results of the discriminant analysis in Study 1.

A one-way analysis of variance (ANOVA) for the three different types of kanji was conducted to examine whether the 18 incorrectly classified kanji of Study 1 were more difficult than other correctly classified kanji with On- and Kun-readings. The results indicated a significant main effect [$F(2,58)=17.66$, $p<.0001$]. Furthermore, a multiple comparison of Tamhame's T2 was conducted for the means of the three different types of kanji. The results indicated that the correct ratio for the 18 incorrectly classified kanji (M = 45.93%) was significantly different from the correct ratio for the 24 kanji with On-readings (M = 70.00%) and the correct ratio for the 19 kanji with Kun-readings (M = 79.30%). These correct ratios for kanji with On-readings and Kun-readings did not differ from each other. Consequently, the kanji items which had been incorrectly classified by discriminant analysis were difficult even for native Japanese speakers to properly identify On- or Kun-readings. The results by humans in Study 2 were likely to resemble the results of the discriminant analysis in Study 1.

Accuracy rates for each of the 18 incorrectly classified kanji are reported in Table 8. Among them, 12 kanji (66.67%) had lower than the random chance level of 50.00 percent. In fact, items with On-readings having a higher semantic concreteness (e.g., 肉, 肺, 液, 菊, 胃, 劇, 服 and 脈) tended to have lower accuracy rate. Native Japanese speakers were likely to misjudge these kanji as Kun-readings. As discussed in the introduction, this tendency is congruent with the

Table 9 Questions for On- or Kun-reading Decision Strategies Based on Kanji Characteristics

#	Kanji Characteristics	Questions of On-/Kun-reading Decision
1	Semantic concreteness	How concrete or abstract is a kanji meaning?
2	Naming latency (ms)	How quickly is a sound of a kanji remembered?
3	Neighborhood size	How many compound words can be constructed by a kanji?
4	Number of kanji homophones	How many kanji share the same sound?
5	Number of morae	How many kana (or morae) make up a sound of a kanji?
6	Special sounds (/N/ or /R/)	Does a kanji contain one of the special sounds, /N/, /Q/ and /R
7	Kanji frequency	How often is a kanji seen?
8	School grade	How early is a kanji learned?
9	Number of strokes	How many strokes are required to write a kanji?
10	Radical frequency	How common is a radical used in a kanji seen?

Note : Shaded variables indicate significant predictors of kanji On- and Kun-readings by discriminant analysis.

Table 10 Percentages of Chosen Reasons for On- or Kun-reading Decision Strategies by 30 Native Japanese Speakers and Category Scores of the Quantification Theory Type III

On- or Kun-reading decision strategies based on Kanji characteristics	% out of 30 subjects	Category scores	
		1st axis	2nd axis
Semantic concreteness	63.33%	-0.131	0.022
Naming latency (ms)	43.33%	0.164	0.032
Neighborhood size	26.67%	-0.149	-0.087
Number of kanji homophones	23.33%	0.073	-0.098
Number of morae	23.33%	0.090	-0.128
Special sounds (/N/ or /R/)	20.00%	0.039	-0.093
Kanji frequency	16.67%	0.067	0.309
School grade	3.33%	-0.142	0.455
Variable Explained (%)		24.4%	20.6%
Accumulative Variable Explained (%)		24.4%	45.0%

Note 1 : Shaded variables indicate strategies chosen by more than 20% of 30 subjects.
Note 2 : Number of strokes and radical frequency were not chosen by any subject, so these two strategies were excluded from the analysis.

proposal by Nomura (1978, 1979) and Kaiho and Nomura (1983). Despite the results of the discriminant analysis, three kanji 酪, 才 and 渓 were correctly identified by participants with an accuracy rate higher than 70 percent. These were low in semantic concreteness, but two kanji, 酪 and 渓, had longer naming latencies. If humans only rely on a single factor of kanji semantic concreteness to distinguish On- and Kun-readings, using the cue of low semantic concreteness, they could be able to identify these kanji correctly as On-readings.

As for Kun-readings, the kanji 瀬 had 701 milliseconds of naming latency which is relatively long among the 18 kanji. This kanji seems to be mistaken as an On-reading by participants. However, the kanji 箱, which obtains a high semantic concreteness, displayed a high correct ratio of 90.00 percent. Although discriminant analysis did not properly identify this kanji, participants could correctly classify it.

8.2.2 Results of Ten Questions of Human Strategies Related to Ten Kanji Characteristics

In Study 2, 10 questions concerning the 10 kanji characteristics were asked to 30 participants after the test. These questions were constructed in such a way as to determine whether they (i.e., native Japanese speakers) actually utilized kanji characteristics to identify On- and Kun-readings. The questions corresponding to these characteristics are shown in Table 9.

The percentages for 30 participants are shown in Table 10 in the order of kanji characteristics chosen in a higher percentage. As expected from

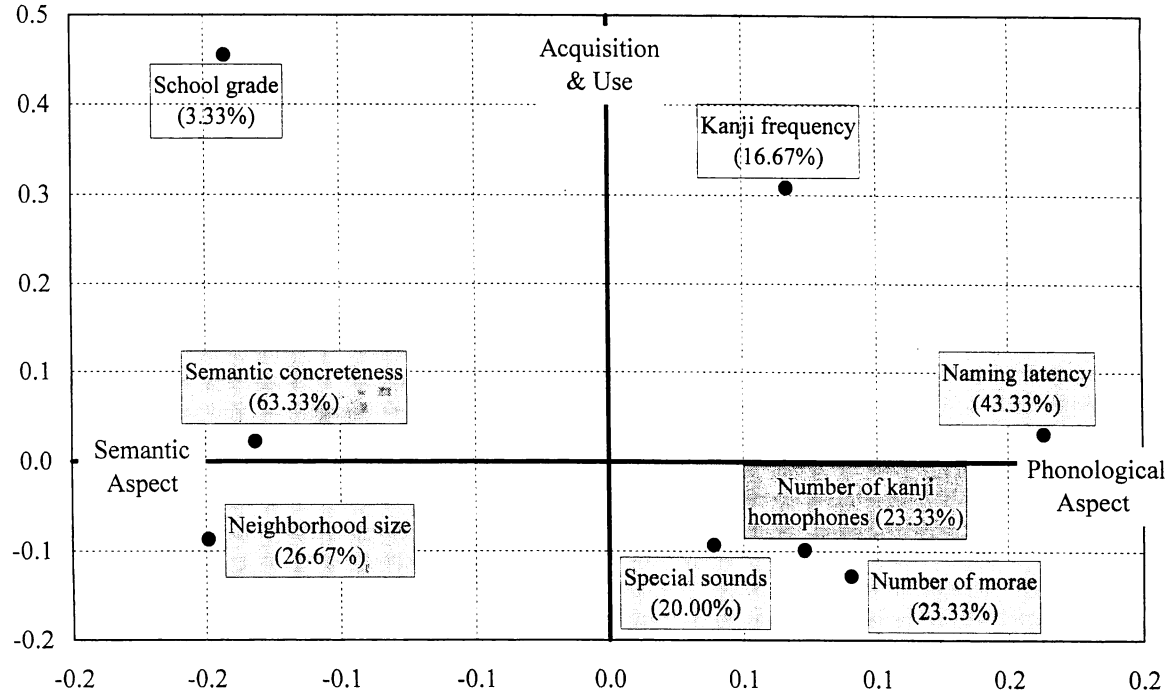

Figure 1 Plotting of On- or Kun-reading decision strategies based on category scores of the quantification theory type III

Note: Shaded variables indicate strategies chosen by more then 20% of 30 subjects.

the results of the discriminant analysis in Study 1, native Japanese speakers also used semantic concreteness (19 out of 30 participants) and naming latency (13 out of 30 participants) as predictors of On- and Kun-readings. A predictor of special sounds for kanji reading was also used as a strategy by Japanese, but was only pointed out by 6 out of 30 participants (20.00%). Number of kanji strokes, which the discriminant analysis indicated as a significant predictor, was not used by participants. Since there was no significant difference in means of kanji strokes between On-reading and Kun-reading in the post-hoc analysis of Study 1, it is reasonable to assume that Japanese participants did not refer to number of strokes to identify kanji readings.

It was rather surprising that native Japanese speakers were aware of using five additional indicators including kanji neighbors (productivity of two-kanji compound words), kanji homophones, constructing morae, kanji frequency and school grade when identifying On- and Kun-readings. Among them, predictors of neighbors, homophones and morae were selected as identifying strategies in more than 6 participants (higher than 20.00%). In fact, the post hoc analysis in Study 1 indicated significant differences between On- and Kun-readings with respect to the following variables: (1) kanji with On-readings had larger kanji neighbors (higher in productivity constructing two-kanji compound words) than kanji with Kun-readings, (2) kanji with On-readings had a larger number of kanji homophones than those with Kun-readings, and (3) kanji with On-readings had fewer morae constructing their pronunciations than those with Kun-readings. These differences support the possibility of reasonable predictors for distinguishing On- and Kun-readings by humans. School grade when kanji were taught was only referred to by a single participant. Finally, radical frequency was not used by participants at all.

In order to classify the 8 strategies selected by 30 Japanese (two strategies, number of strokes and radical frequency, were not selected by any of the participants), the quantification theory type III analysis (the same as correspondence analysis for free-response type data) was conducted. Category scores of the first and second axes were reported in Table 10 and plotted in Figure 1. As shown in Figure 1, there are three groups of human strategies on the horizontal axis from the semantic to phonological aspects. The first group is semantic concreteness and kanji neighborhood size. Since 'kanji neighbor' refers to how often the characters produce compound words, this variable might be strongly related to semantic aspects. The second group includes only one strategy, naming latency, indicating how fast Japanese can remember a kanji sound. The third group, which is related to the phonological aspect, included number of morae, number of kanji homophones and special sounds. These characteristics were not chosen by more than 6 (or 20%) of the 30 participants. These results imply that some participants actually look at phonological structure and features of kanji sounds in order to determine On- or Kun-readings. The two strategies of school grade and kanji frequency were not so often used for On- or Kun-reading decision making, and as shown in Figure 1, they are not related each other.

8.3 Discussion of Study 2

Study 2 focused on the human strategies to identify On- and Kun-readings. First, an On- or Kun-reading test was administered to investigate whether or not native Japanese speakers have difficulty in properly identifying On- and Kun-readings of 18 kanji which were incorrectly classified by the discriminant analysis in Study 1. Three kanji groups of 18 incorrectly classified kanji, 19 with Kun-readings and 24 with On-readings displayed similar trends in accuracy as shown by discriminant analysis. In other words, native Japanese speakers generally exhibit similar accuracy tendencies as the discriminant analysis. The test to identify kanji readings indicated that 45.93 percent of the readings of these 18 kanji were properly identified. This rate of accuracy is slightly lower than the random chance level of 50.00 percent when one without knowledge of kanji chooses one of two possible readings randomly. The kanji items misclassified by the discriminant analysis were also difficult to identify for native Japanese speakers. However, a detailed inspection of the incorrectly classified 18 kanji revealed some differences in accuracy between discriminant analysis and participants. The readings of four kanji, 酪, 才, 渓 and 箱 were properly identified at a rate higher than 70.00 percent by participants. The questionnaire used to identify On- and Kun-reading, asked all participants whether or not each of the 10 kanji characteristics was utilized. Kanji characteristics of semantic concreteness and naming latency were major predictors for identifying readings used by native Japanese speakers. The three groups of On- or Kun decision strategies were identified by quantification theory analysis; (1) semantic concreteness and neighborhood size, (2) naming latency, and (3) number of kanji homophones, number of morae and special sounds. These results suggest that participants rely on multiple kanji characteristics. Study 2 also indicated that native Japanese speakers do not use kanji strokes and radical frequency for On and Kun identification.

9. General Discussion — Congruency between Indicators chosen by Discriminant Analysis and Strategies Selected by Humans

The main purpose of the present study was to establish links between kanji characteristics and human strategies when distinguishing On- and Kun-readings from one another. Study 1 asked which kanji characteristics are significant indicators of On- and Kun-readings for kanji. Discriminant analysis was conducted to find significant indicators among the 10 selected characteristics, resulting in four significant indicators being found (see Table 6). Due mostly to these four indicators, 116 kanji of 134 (86.57%) were properly classified into On- or Kun-readings (see

Table 5).

Despite the high accuracy rate of classification by discriminant analysis, there were still 18 kanji items (see Table 8) which were not properly classified. Thus, Study 2 tested the extent to which native Japanese speakers could properly identify On- and Kun-readings for these 18 kanji. The accuracy rate of 30 native Japanese speakers was only 45.93 percent for these 18 kanji (see Table 7), less than the random chance level of 50 percent. Thus, even native Japanese speakers had great difficulty to correctly identify On- and Kun-readings for these 18 kanji. In addition, after the test, a questionnaire concerning the use of the 10 kanji characteristics for identifying On- or Kun-readings was given to the 30 participants (see Table 9). The questionnaire study revealed three congruent indicators/strategies between discriminant analysis and participants (i.e., semantic concreteness, naming latency, special sounds). One of the four significant indicators found in the discriminant analysis was hardly used by participants (i.e., number of strokes) as a strategy. Although native Japanese were likely to follow the trend as displayed by discriminant analysis, the questionnaire uncovered a variety of strategies used by native Japanese speakers to identify On- and Kun-readings. Specifics as to congruence and incongruence in indicators/strategies between discriminant analysis and human strategies are discussed in the following sections.

9.1 Congruent Kanji Characteristics as Indicators and Strategies

Three congruent indicators/strategies were found between the discriminant analysis in Study 1 and human strategies identified in Study 2: Semantic concreteness, naming latency and special sounds.

9.1.1 Semantic Concreteness

When Chinese characters were adopted to represent *wago* and their meanings, Kun-readings were added to the characters to allow words to keep their traditional Japanese sounds. Takashima (2001) provides an interesting explanation. When the kanji 山 was adopted into the Japanese language, it was sounded by an On-reading /saN/. Since Japanese understood its kanji meaning 'mountains', they attached a Kun-reading /yama/ which had already existed in the Japanese vocabulary. Takashima considered this process a novel idea, almost as if the English word 'dog' was to be sounded /inu/ in Japanese as well as in English. This provides an explanation of why kanji with Kun-readings are likely to represent concrete meanings such as 'waterfall' (滝 /taki/), 'shellfish' (貝 /kai/), 'daughter' (娘 /musume/) and 'box' (箱 /hako/). In accordance with this explanation, discriminant analysis in Study 1 revealed that degree of semantic concreteness was the most significant indicator of On- and Kun-readings. Native Japanese speakers were also aware of this kanji characteristic indicating Kun-readings, and used it to identify On- and Kun-readings. This finding supported previous studies of Hirose (1998), Nomura (1978, 1979) and Kaiho and Nomura (1983).

9.1.2 Naming Latency

The discriminant analysis in Study 1 showed that naming latency, which indicates how quickly one can find a kanji pronunciation, was the second most significant indicator of the ten variables. Kanji with On-readings are very often found in various compound words, classified as *kango*. The meaning of *kango* is not solely based upon one particular kanji in the compound structure but on all kanji therein. Thus, the meanings of kanji with On-readings vary depending upon how they are combined with other kanji to create *kango*. For example, a kanji 芸 creates various two-kanji compound *kango* words including 芸人 /geiniN/ 'performer', 芸当 /geitoR/ 'trick', 芸名 /geimei/ 'screen name', and 無芸 /mugei/ 'uncultured person' (examples taken from Nelson, 1992). As a result of this particular feature, the meanings of single kanji with On-readings have become ambiguous, and therefore are related more closely to phonology than semantics. As found in the discriminant analysis, native Japanese speakers also indicated quick retrieval of sounds as an important strategy to determine On- or Kun-readings. In fact, psychological studies by Hirose (1998), Nomura (1978, 1979) and Kaiho and Nomura (1983) indicate that On-readings are more strongly attached to kanji phonology while Kun-readings are more attached to semantics.

9.1.3 Special Sounds

The present investigation as to whether or not special Japanese sounds constitute a significant indicator of kanji reading is unique since previous studies by Hirose (1998), Nomura (1978, 1979) and Kaiho and Nomura (1983) did not consider them as a useful indicator of On-readings used by native Japanese speakers. Historically, the three special sounds of /N/, /Q/ and /R/ were introduced into Japanese phonology with the adoption of Chinese characters in the Heian Period (Komatsu, 1981; Kubozono & Ota, 1998; Numoto, 1987). Thus, from a linguistic perceptive, these sounds are frequently associated with On-readings. Tamaoka et al. (2001, 2002) indicated that half of the 1,945 basic kanji with On-readings contain these special sounds (see Table 2), whereas they are contained in only 48 of the 1,945 (3.94%) kanji with Kun-readings (see Table 3). As was expected, special sounds were a significant indicator of kanji reading, ranking third out of the ten variables tested by discriminant analysis in Study 1. Furthermore, Study 2 directly asked participants whether they utilized special sounds to find On- and Kun-readings. Six of 30 native Japanese speakers were aware of making use of special sounds for distinguishing between the two readings.

It has been argued at length whether or not native Japanese speakers process Japanese on the basis of mora or syllable (e.g., Otake, 1996; Cutler & Otake, 1994; Otake, Hatano, Chutler & Mehler, 1993; Tamaoka & Terao, 1999). Within this topic, special sounds are found never to stand alone, even though they are considered as a single mora. Special sounds are always combined with another mora to form a single syllable. For example, in the sound of /kiN/, although /ki/ counts as one mora and /N/ as another, there is only one syllable in this sound. In this sense, as proposed by Tamaoka and Terao (1999), it is quite possible that native Japanese speakers perceive and produce kanji sounds such as /kiN/, /gaN/ and /soR/ as a single unit on the basis of syllable. Native Japanese speakers must be sensitive to these special sounds in syllabic units, and some are likely Japanese to use these sounds as a strategy to identify On-readings.

9.2 Non-Significant Indicators Selected as Strategies

A test identifying On- or Kun-readings was administered to 30 native Japanese speakers in Study 2. The results displayed a similar tendency as the accuracy rates of the discriminant analysis. However, the questionnaire following the test provided evidence that native Japanese speakers were aware of various strategies which the discriminant analysis did not reveal. Five different strategies were taken as cues to identify On- and Kun-readings.

9.2.1 Kanji Neighborhood size

As discussed in the introduction, a majority of kanji compound words with On-readings were created for the purposes of translating various new words from books written in alphabetic scripts in the late Edo period and in the Meiji period (Kabashima, 1989, Takashima, 2001). Thus, in the present study, kanji neighborhood size, which was defined as the number of times one kanji can combine with another to create two-kanji compound words (Kawakami, 1997; Tamaoka et al., 2001, 2002), was chosen as one of the ten kanji characteristics which contribute to the ability to distinguish between On- and Kun-readings. However, kanji neighborhood size proved not to be a significant indicator of kanji reading by the discriminant analysis. Despite the results of Study 1, Study 2 suggested that native Japanese speakers (at least 8 of 30 participants) indicated their use of this kanji characteristic. Indeed, the post-hoc analysis in Study 1 showed that kanji with On-readings had a greater kanji neighborhood size than those with Kun-readings; it is therefore possible that native Japanese speakers use an approach whereby if they can easily find some compound words produced by a target kanji, then they can judge this kanji as On-reading.

9.2.2 Kanji homophony

A majority of kanji homophones are found among kanji with On-readings; the 113 selected kanji with On-readings had a little more than 15 homophones on average (see Appendix) while the 21 selected kanji with Kun-readings had no kanji homophones (i.e., had a single sound per kanji). Therefore, the present study assumed that kanji homophony would be a strong candidate for distinguishing between On- and Kun-readings. However, this was shown not to be the case by the results of the discriminant analysis in Study 1.

Notwithstanding this result, the questionnaire in Study 2 showed that 7 out of the 30 native Japanese speakers checked kanji homophony as a strategy to distinguish between On- and Kun-readings. In applying this strategy, they must recall some homophonic kanji sharing the same sound. For example, from a target kanji 閣 /kaku/, they must retrieve various kanji of the same sound such as 画, 核, 格, 各 and 角. Then, when recalling some kanji sharing the same sound, they can determine that the target kanji 閣 /kaku/ is an On-reading.

The average number of kanji homophones among On-readings was much greater than Kun-readings. However, still 9 out of the 113 selected On-reading kanji (i.e., 肉 /niku/, 域 /iki/, 菊 /kiku/, 寸 /suN/, 税 /zei/, 脈 /myaku/, 陸 /riku/, 軸 /ziku/, and 百 /hyaku/) had no kanji homophones (see the Appendix). It seems that the number of kanji homophones varies depending upon the character in question. As a result, kanji homophony was not shown to be a significant indicator of kanji reading by the discriminant analysis in Study 1. Yet, in that some kanji with On-readings indeed have a great number of kanji homophones, it is assumed that native Japanese speakers may selectively use kanji homophony as a strategy to identify On-readings.

9.2.3 Number of Morae

Although the discriminant analysis did not select the number of morae as a significant indicator, Study 2 confirmed that 7 out of 30 participants used it as a strategy to identify On- and Kun-readings. Kanji with a large number of morae are often found in Kun-readings such as 湖, sounded /mizu'umi/, which has four morae. The 113 On-reading kanji used in the present study consisted of no more than two morae each. Three kanji out of the 21 Kun-reading kanji had three morae, namely 娘 /musume/, 岬 /misaki/ and 峠 /toRge/. Native Japanese speakers may have used this strategy for these kanji. However, because no other Kun-readings of kanji consisted of more than two morae, it seems these three kanji were the only ones to which this strategy can be actually applied.

9.2.4 Kanji frequency

Although kanji frequency is considered a strong influence on cognitive processing (e.g., Fushimi, Ijuin, Patterson & Tatsumi, 1999; Hino & Lupker, 1998; Tamaoka & Takahashi, 1999), according to the results of the discriminant analysis in Study 1, kanji frequency was the weakest none-significant indicator among the ten characteristics of kanji. Because the means of kanji frequency between those with On- and Kun-readings did not differ significantly, this result is quite understandable; kanji frequency is not directly related to the phonological aspect of On- and Kun-readings. Despite this finding, 5 out of 30 participants indicated they used kanji frequency as a strategy to identify On- and Kun-readings. Since kanji with Kun-readings can be often seen in print as a single kanji word, they may identify these kanji as Kun-readings.

9.2.5 School Grade

School grade, used an index for one's age of kanji acquisition, indicated strong significant correlations with neighborhood size, naming latency and kanji frequency (see Table 4). However, the discriminant analysis in Study 1 did not identify school grade as significant for distinguishing between On- and Kun-readings. Only one participant out of 30 in Study 2 cited use of school grade as a strategy. In fact, 38 out of 80 kanji (47.5%) in the first grade are pictographic kanji such as 雨 /ame/ ('rain'), 犬 /inu/ ('dog'), 耳 /mimi/ ('ears'), 車/kuruma/ ('car'), and 水 /mizu/ ('water'). These pictographic kanji stand alone as single words being read in Kun-readings. In the second grade, this proportion decreased to 55 pictographic kanji out of 160 or 34.38 percent. In the third grade the proportion becomes 24 out of 200 kanji or 12.00 percent. Since pictographic kanji only comprise 12.85 percent of the 1,945 basic kanji (Tamaoka, et al., 2001, 2002), the proportion of pictographic kanji is higher in the first and second grades. Thus, it is possible to apply this strategy for some kanji taught in early grades, but its usage is quite limited and not so accurate. It was perhaps due to these reasons that only one participant checked this strategy.

9.3 Significant Indicator not Selected as Strategy

According to the results of Study 1, number of strokes was found to be an unexpected significant indicator of kanji reading. As number of strokes is a characteristic related to orthography, rather than phonology, it had not been expected to rank as the fourth most significant indicator. Kanji with On-readings had 1.24 strokes more than those with Kun-readings. However, this difference was not significant when comparing the means of strokes between kanji with On- and Kun-readings. This small difference does seem to contribute to distinguishing between On- and Kun-readings of kanji. In contrast, number of strokes was not selected by native Japanese speakers as a strategy to determine On- and Kun-readings. Because there is not significant difference in strokes between kanji with On- and Kun-readings, it is assumed that native Japanese speakers did not use strokes representing a degree of visual complexity as a reliable indicator.

9.4 Kanji Characteristics of Neither Indicator nor Strategy

Radical frequency refers to the number of kanji which share the same radical from among the 1,945 basic Japanese kanji. For example, *sanzui*, the most common kanji radical meaning 'water,' is used as a constructing component in 103 of the 1,945 basic kanji (Tamaoka & Yamada, 2000; Tamaoka, et al., 2001, 2002). The top ten most frequently-used radicals are utilized to construct about 34 percent (i.e., 669 kanji) of the 1,945 basic kanji. This tendency is also true among Chinese characters used in the Chinese language: 17 radicals out of 214 are used as basic elements in constructing 50.17 percent of 8,711 Chinese characters (Leong, 1973). Previous studies (e.g., Leong & Tamaoka, 1995; Saito, 1997; Saito, Masuda & Kawakami, 1998) indicate that radicals serve as a fundamental element of kanji which have some effects on the cognitive processing of a whole kanji. However, the index of radical frequency did not appear to be an effective indicator/strategy of On- and Kun-readings as determined by either discriminant analysis in Study 1 or human participants in Study 2.

Study 1: Indicators Discriminant Analysis	Study 2: Strategies Human Participants
1 Semantic concreteness	1 Semantic concreteness
2 Naming latency	2 Naming latency
3 Special sounds (/N/ or /R/)	3 Neighborhood size
4 Number of strokes	4 Number of kanji homophones
5 Radical frequency	5 Number of morae
6 Number of kanji homophones	6 Special sounds (/N/ or /R/)
7 School grade	7 Kanji frequency
8 Number of morae	8 School grade
9 Neighborhood size	9 Number of strokes
10 Kanji frequency	10 Radical frequency

Figure 2 Comparison of Indicators and Strategies for Identifying On- and Kun-readings
Note 1: Shaded variables in Study 1 were significant indicators chosen by discriminant analysis.
Note 2: Shaded variables in Study 2 were selected by more than 6 subjects (20%) of 30 subjects.

9.5 Summary — Comparison of Statistically Predicted Indicators and Human Strategies

The present study investigated the issue of identifying attributes of kanji On- and Kun-readings from the perspectives of both statistical prediction and human selection. As shown in Figure 2, both approaches brought to the fore some kanji characteristics which contributed to selection of On- and Kun-reading.

In Study 1, a discriminant analysis using the stepwise method provided the best of four significant indicators for distinguishing On- and Kun-readings of the 134 kanji. These are semantic concreteness, naming latency (i.e., speed of sound retrieval), special sound and number of strokes. Since these indicators were selected by using the purely quantified data of kanji characteristics in the approach of predicting On- and Kun-readings, the results of the discriminant analysis are mathematically sound. However, the results do not guarantee whether native Japanese speakers actually use these significant indicators as strategies to identify On- and Kun-readings.

In Study 2, the identifying test of On- and Kun-reading displayed a similar tendency to the accuracy rates of discriminant analysis. After the test, a questionnaire was administered whereby six out of 10 strategies used by participants to identify On- and Kun-readings were checked by more than 6 out of the 30 native Japanese speakers. Three of these six strategies were congruent with significant indicators specified by discriminant analysis, namely, semantic concreteness, naming latency and special sounds. Despite the significant indicator in Study 1, the strategy concerning kanji strokes was not used by humans. Also, humans are more likely to use neighborhood size (i.e., number of kanji compound words produced by a target kanji), kanji homophones and phonological structure (i.e., number of morae).

The quantification analysis (see Figure 1) showed three categories of strategies used by participants. Semantic concreteness and kanji neighborhood size were classified together from the semantic aspect. Native Japanese speakers are likely to check whether a kanji contains a concrete meaning when identifying On- or Kun-reading. Free-standing lexical kanji are very often read in Kun-reading. If a kanji cannot be readily identified, a native Japanese speaker may try to produce a two-kanji compound word which is frequently read in On-reading. Kanji neighborhood size refers to the number of kanji compound words produced by a single kanji, so this indicator can accompany semantic concreteness. The second category is a single indicator — naming latency from the phonological aspect. Considering Nomura's research (1978, *1979) which in-*

dicated that On-readings are pronounced more quickly than Kun-readings, the present study suggests that Japanese may use accessibility of sounds as a reference. The third group included three indicators which are all related to kanji sounds: number of morae, number of kanji homophones and special sounds. The structure of kanji pronunciation may aid native Japanese speakers somewhat in determining kanji readings. It is likely that humans use all these strategies simultaneously in order to identify kanji On- and Kun-readings.

The present study showed discrepancies between statistical analysis and human strategies in determining On- and Kun-reading. Discriminant analysis indicates only rule-based judgments on the basis of kanji characteristics. In contrast, humans are likely to use a wider variety of strategies when distinguishing between On- and Kun-readings. It is indeed inefficient to use kanji characteristics which do not differ among kanji when making decisions as to reading. For example, kanji such as 肉 /niku/ ('meat') and 菊 /kiku/ ('chrysanthemum') are free-standing, clear and meaningful units by themselves; they also do not contain special sounds. On the basis of discriminant analysis, these kanji would be judged as Kun-readings even though they are in fact On-readings. As such, the degree of differences indicated by kanji characteristics changes from one kanji to another. The results between indicators and strategies illustrate a more general point: On- and Kun-readings can be effectively predicted by discriminant analysis on the basis of various kanji characteristics; however, due to a lack of consistency in On- and Kun-readings attached to each kanji, humans can flexibly incorporate a wider variety of strategies when making their determinations.

References

Balota, D. A. & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, **10**, 340–357.

Coltheart, M. (1985). Cognitive neuropsychology and the study of reading. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI*, 3–37. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Coltheart, M. (1987). Functional architecture of the language-processing system. In M. Coltheart, G. Sartori & R. Job (Eds.), *The cognitive neuropsychology of language*, 1–25. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance IV*, 535–555. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed processing approaches. *Psychological Review*, **100**, 589–608.

Cutler, A. & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, **33**, 824–844.

Forster, K. I. (1979). Levels of processing and the structure of the language processes. In W. E. Cooper and E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merril Garret*, 27–84. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fushimi, T., Ijuin, M., Patterson, K., & Tatsumi, I. F. (1999). Consistency, frequency, and lexicality effects in naming Japanese kanji. *Journal of Experimental Psychology: Human Perception and Performance*, **25**, 382–407.

Goryo, K. (1987). *Yomu to yuukoto [Concerning Reading]*. Tokyo: Tokyo University Press. (In Japanese)

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Languages*, **29**, 228–244.

Hadamitzky, W. & Spahn, M. (1981). *Kanji and kana: A handbook and dictionary of the Japanese writing system.* Tokyo: Charles E. Tuttle Company.

Hino, Y. & Lupker, S. L. (1998). Effects of word frequency for Japanese kana and kanji words naming and lexical decision: Can the dual-route model save the lexical-selection account? *Journal of Experimental Psychology: Human Perception and Performance*, **24**, 1431–1453.

Hirose, H. (1998). Identifying the On- and Kun-readings of Chinese characters: Identification of On versus Kun as a strategy-based judgment. In Leong, C.K. & Tamaoka, K. (Eds.), *Cognitive processing of the Chinese and the Japanese languages*, 375–394. Dordrecht, the Netherlands: Kluwer Academic Publishers.

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

Kaiho, H. & Nomura, Y. (1983). *Kanji joho shori no shinrigaku [The psychology of kanji information processing]*. Tokyo: Kyoiku Shuppan. (In Japanese)

Kamata, T. (Ed.). (1991). *Kuwashii shoogakkoo kanji jiten [A comprehensive elementary school kanji dictionary]*. Tokyo: Bun'eido. (In Japanese)

Kato, M. (1989). Jôyô kanji no jisû, jishu, jitai [Numbers, types and styles of commonly-used kanji]. In K. Satô (Ed.), *Kanji to kokugo mondai [Kanji and its problems related to the Japanese language]*, 26–37. Tokyo: Meiji Shoin. (In Japanese)

Kawakami, M. (1997). JIS 1-shu kanji 2,965 ji wo mochiite sakusei sareru kanji niji jukugoshu hyo [Tables of two-kanji compound words constructed by 2,965 JIS 1st kanji characters]. *Kyoiku Gakubu Kiyo [School of Education Bulletin]*, Nagoya University (Psychology), **44**, 243–299. (In Japanese)

Kayamoto, Y., Yamada, J., & Takashima, H. (1998). The consistency of multiple-pronunciation effects in reading: The case of Japanese logographs. *Journal of Psycholinguistic Research*, **27**, 619–637.

Kayamoto, Y. (2000). Nihongo wo gakushû suru chugokugo bogo washa no kanji ninchi — Jokyu/chojokyu no naiteki jisho ni okeru on'in joho shori [Processing phonological information: Recognition of Japanese characters by advanced- and superior-level native speakers of Chinese]. *Japanese Journal of Educational Psychology*, **48**, 315–322. (In Japanese)

Kess, J. F. & Miyamoto, T. (1999). *The Japanese mental lexicon: Psycholinguistic studies of kana and kanji processing*. Philadelphia, PA: John Benjamins Publishing Company.

Kobayashi, Y. (1998). *Nihon no kanji [Japanese kanji]*. Tokyo: Taishukan Shoten.

Kokuritsu Kokugo Kenkyujo [National Institute for Japanese Language] (1962). *Gendai zasshi 90 shu no yogo yorei (Vol. 2) [Word usage in 90 current Japanese magazines]*. Tokyo: Shuei Shuppan. (In Japanese)

Komatsu, H. (1981). *Nihongo no on'in [Japanese phonology]*. Tokyo: Chuo Koronsha. (In Japanese)

Kubozono, H. & Ota, S. (1998). *On'in kozo to akusento [Phonological structure and accents]*. Tokyo: Kenkyusha Shuppan. (In Japanese)

Leong, C. K. (1973). Hong Kong. In J. Downing (Ed.), *Comparative reading: Cross-national studies of behavior and process in reading and writing*, 383–402. New York: Macmillan Company.

Leong, C. K. & Tamaoka, K. (1995). Use of phonological information in processing kanji and katakana by skilled and less skilled Japanese readers. *Reading and Writing*, **7**, 377–393.

Miller, R. A. (1967). *The Japanese language*. Chicago: The University of Chicago Press.

Ministry of Education, Culture, Science, Sports and Technology, Government of Japan (1978). *Chugakko shidosho — Kokugohen [The Japanese language — the course of study at junior high school]*. Tokyo: Tokyo Shoseki. (In Japanese)

Ministry of Education, Culture, Science, Sports and Technology, Government of Japan (1987). *Shogakko shidosho — Kokugohen [The Japanese language — the course of study at elementary school]*. Osaka: Osaka Shoseki. (In Japanese)

Ministry of Education, Culture, Science, Sports and Technology, Government of Japan (1998). *Monbusho kokuji — Shogakko shido yoryo [The announcement of the elementary school course of study by the Ministry of Education, Science and Culture, Government of Japan]*. Tokyo: Gyosei. (In Japanese)

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, **118**, 43–71.

Nelson, A. N. (1992). The Modern Reader's Japanese-English Character Dictionary (35th print). Tokyo: Charles E. Tuttle.

Nomura, K. (1988). *Kanji no mirai [The future of kanji]*. Tokyo: Chikuma Shobo. (In Japanese)

Nomura, Y. (1978). Kanji no joho shori: Ondoku kun-doku no imi no fuka [The information processing of Chinese characters (kanji): Chinese reading, Japanese reading and the attachment of meaning]. *Japanese Journal of Psychology*, **49**, 190–197. (In Japanese)

Nomura, Y. (1979). Kanji no joho shori: Ondoku kun-doku no kensaku katei [Information processing of Chinese characters (Kanji): Retrieval processes in Chinese-style reading (On) and Japanese-style reading (Kun)]. *Japanese Journal of Psychology*, **50**, 101–105. (In Japanese)

Numoto, K. (1987). *Nihon kanjion no kenkyu [A study on Japanese kanji sounds]*. Tokyo: Toyodo Shuppan. (In Japanese)

Otake, T. (1996). Onsei no chikaku [Sound Recognition]. In Y. Otsu (Ed.), *Ninchi shinrigaku 3 — Gengo [Cognitive psychology 3: Language]*, 131–144. Tokyo: Tokyo Daigaku Shuppankai. (In Japanese)

Otake, T., Hatano, G., Chutler, A. & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, **32**, 258–278.

Sato K. ed. (1996). *Kanji hyakka dai jiten [Kanji Encyclopedia]*. Tokyo: Meiji Shoin.

Saito, H. (1981). Kanji to kana no yomi ni okeru keitaiteki fugooka oyobi on'inteki fugooka no kentoo [Use of graphemic and phonemic encoding in reading Kanji and Kana]. *Japanese Journal of Psychology*, **52**, 266–273. (In Japanese)

Saito, H. (1997). *Shinteki jisho [The mental lexicon]*. In Y. Otsu & T. Gunji (Eds.), *Linguistic Sciences, Vol. 3*, 93–153. Tokyo: Iwanami Shoten. (In Japanese)

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*, 169–203. London: Kluwer Academic Publishers.

Seeley, C. (1984a). Aspects of the Japanese language writing system. *Visible Language*, **18**, 213–218.

Seeley, C. (1984b). The Japanese script since 1900. *Visible Language*, **18**, 267–301.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, **7**, 263–272.

Taft, M. (1991). *Reading and the mental lexicon.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Takamatsu, M. (1997). *Nihon kanji on'inron [Phonology of Japanese kanji]*. Tokyo: Kazama Shobo. (In Japanese)

Takashima, T. (2001). *Kanji to nihon'jin [Kanji and Japanese]*. Tokyko: Bungei Shunju. (In Japanese)

Tamaoka, K. (1991). Psycholinguistic nature of the Japanese orthography. *Gengo Bunka Kenkyu [Studies in Language and Literature] (Matsuyama University)*, **11** (1), 49–82.

Tamaoka, K. (2002). A phonological shift of Japanese kanji with 50 ± 5 percent On-reading ratio embedded in a high On- or Kun-reading environment. *The Proceedings of the 19th Annual Meeting of the Japanese Cognitive Science Society*, 68–69, Japan Advanced Institute of Science and Technology, June 14-15, 2002.

Tamaoka, K. & Hatsuzuka, M. (1995). Kanji niji jukugo no shori ni okeru kanji shiyo hindo no eikyo [The effects of kanji printed-frequency on the processing of Japanese two-morpheme compound words]. *The Science of Reading*, **39**, 121–137. (in Japanese with a substantial English summary)

Tamaoka, T. & Takahashi, N. (1999). Kanji niji jyukugo no shoji koodoo ni okeru goi shiyoo hindo oyobi shojiteki fukuzatsusei 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. (In Japanese)

Tamaoka, K. & Terao, Y. (1999). The phonological processing unit used by native Japanese speakers. Program abstracts, the 12th World

Congress of Applied Linguistics (Waseda University, Japan), 388.

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.

Tamaoka, K., Kirsner, K., Yanase, Y.,Miyaoka, Y., & Kawakami, M. (2001). *Database for the 1,945 basic Japanese kanji (2nd Edition)*. Oxford Text Archive, Oxford University, U.K. The downloadable kanji database (2003, 3rd Edition) is also available at the homepage of www.iie.hiroshima-u.ac.jp/center/tamaoka/down.htm.

Tamaoka, K., Kirsner, K., Yanase, Y.,Miyaoka, Y., & Kawakami, M. (2002). A web-accessible database of the 1,945 basic Japanese kanji. *Behavioral Research Methods, Instruments, & Computers*, **34**, 260–275.

Tanaka, A. (1978). *Kokugo goi ron [A study of the Japanese vocabulary]*. Tokyo: Tokyo Shoin. (In Japanese)

Tsukishima, H. (1979). *Kokugogaku [A study of the Japanese language]*. Tokyo: Tokyo Daigaku Shuppankai [Tokyo University Press]. (In Japanese)

Watanabe, M. (1989). Joyo kanji no onkun [On- and Kun-readings of the commonly-used kanji]. In K. Sato (Ed.), *Kanji to kokugo mondai [Kanji and its problems related to the Japanese language]*, 38–64. Tokyo: Meiji Shoin. (In Japanese)

Whaley, C. P. (1978). Word — nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, **17**, 143–154.

Wydell, T. N., Butterworth, B., & Patterson, K. E. (1995). The inconsistency of consistency effects in reading: The case of Japanese kanji. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 1155–1168.

Yokoyama, S., Sasahara, H., Nozaki, H., & Long, E. (1998). *Shinbun denshi media no kanji — Asahi shinbun CD-ROM niyoru kanji hindo hyo [Japanese kanji in the newspaper media — Kanji frequency index from the Asahi Newspaper on CD-ROM]*. Tokyo: Sanseido. (In Japanese)

(Received 7 May 2001)
(Accepted 21 March 2003)

Katsuo Tamaoka

Katsuo Tamaoka is Professor at the Institute for International Education, Hiroshima University, Japan. He received his Ph. D. in Psychology from the University of Saskatchewan, Canada. His research involves the construction of cognitive models for languages, by experimenting with the effects of phonological, morphological, lexical and syntactic structures on the cognitive processing of Japanese. His studies also extend to corpus studies for analyzing linguistic features and to continuous evaluation for language acquisition.

Appendix Kanji Stimulus Items Used for Study 1 (n = 134)

Kanji	Readings	Phonemic Sound	Special Sounds	Naming Latency (ms)	Kanji Frequency	Radical Frequency	Neighborhood Size	Semantic Concreteness	Number of Homophones	Number of Strokes	Number of Morae	School Grade
劇	On	/geki/	-	611	5912	24	59	5.98	3	15	2	6
族	On	/zoku/	-	585	9532	5	70	4.52	5	11	2	3
段	On	/daN/	/N/	558	7993	5	64	4.15	8	9	2	6
題	On	/dai/	-	573	32315	17	86	4.67	5	18	2	3
脳	On	/noR/	/R/	523	6385	28	38	6.17	6	11	2	6
糖	On	/toR/	/R/	577	725	10	24	4.71	37	16	2	6
線	On	/seN/	/N/	606	11809	52	120	5.35	26	15	2	2
芸	On	/gei/	-	543	5005	38	75	5.04	3	7	2	4
案	On	/aN/	/N/	580	1907	23	70	4.02	3	10	2	4
県	On	/geN/	/N/	616	29515	15	32	5.33	29	9	2	3
軍	On	/guN/	/N/	520	16031	4	165	5.45	3	9	2	4
式	On	/siki/	-	574	15274	2	97	4.23	2	6	2	3
点	On	/teN/	/N/	542	26969	10	165	4.77	7	9	2	2
肉	On	/niku/	-	549	2707	2	100	5.88	1	6	2	2
湾	On	/waN/	/N/	613	4362	103	15	5.23	2	12	2	7
胃	On	/i/	-	557	352	11	18	6.23	21	9	1	4
徳	On	/toku/	-	573	3572	19	149	3.58	6	14	2	5
域	On	/iki/	-	583	10437	25	38	5.06	1	11	2	6
塁	On	/rui/	-	630	5303	20	29	3.96	4	12	2	7
液	On	/eki/	-	619	1550	103	41	5.33	5	11	2	5
億	On	/oku/	-	547	18874	87	6	4.42	3	15	2	4
可	On	/ka/	-	547	11946	38	41	3.56	31	5	1	5
服	On	/huku/	-	599	3759	28	114	5.92	8	8	2	3
王	On	/oR/	/R/	546	5126	9	133	5.31	12	4	2	1
科	On	/ka/	-	550	8197	20	78	3.04	31	9	1	2
閣	On	/kaku/	-	763	8308	12	35	3.25	16	14	2	6
刊	On	/kaN/	/N/	604	3482	24	28	3.25	44	5	2	5
課	On	/ka/	-	563	8353	60	31	3.71	31	15	1	4
汽	On	/ki/	-	612	156	103	11	3.08	34	7	1	2
艦	On	/kaN/	/N/	712	808	6	31	4.13	44	21	2	7
菌	On	/kiN/	/N/	751	860	38	24	5.56	12	11	2	7
具	On	/gu/	-	567	6043	8	110	4.23	2	8	1	3
郡	On	/guN/	/N/	648	2100	11	21	4.52	3	10	2	4
閑	On	/kaN/	/N/	962	138	12	57	3.29	44	12	2	7
渓	On	/kei/	-	1011	108	103	17	3.67	26	11	2	7
隊	On	/tai/	-	584	8901	28	36	4.85	17	12	2	4
玄	On	/geN/	/N/	653	700	2	62	3.27	10	5	2	7
碁	On	/go/	-	744	284	3	26	4.79	11	13	1	7
昆	On	/koN/	/N/	748	249	20	8	3.46	11	8	2	7
菊	On	/kiku/	-	623	1050	38	66	6.29	1	11	2	7
券	On	/keN/	/N/	594	6734	5	27	5.60	29	8	2	5
愛	On	/ai/	-	533	6943	40	107	4.88	2	13	2	4
剤	On	/zai/	-	678	2064	24	36	4.04	5	10	2	7
郊	On	/koR/	/R/	665	1053	11	15	3.46	60	9	2	7
策	On	/saku/	-	617	20888	21	55	4.29	8	12	2	6
史	On	/si/	-	572	8119	38	87	3.90	45	5	1	4
校	On	/koR/	/R/	629	18285	54	74	3.85	60	10	2	1
詩	On	/si/	-	595	1495	60	110	5.69	45	13	1	3
週	On	/syuR/	/R/	597	5794	50	19	4.88	21	11	2	2
銃	On	/zuR/	/R/	579	2092	28	39	5.96	11	14	2	7
順	On	/zuN/	/N/	539	4489	17	74	4.29	11	12	2	4
署	On	/syo/	-	635	414	15	21	4.08	9	13	1	6
症	On	/shoR/	/R/	585	2306	12	21	4.04	49	10	2	7
誌	On	/si/	-	560	3336	60	17	4.17	45	14	1	6
陣	On	/ziN/	/N/	580	3614	28	79	4.58	8	10	2	7
睡	On	/sui/	-	740	162	5	21	4.02	13	13	2	7
章	On	/syoR/	/R/	559	3010	5	59	4.23	49	11	2	3
寸	On	/suN/	/N/	580	363	12	68	4.17	1	3	2	6
聖	On	/sei/	-	597	1184	2	137	3.75	31	13	2	6
仙	On	/seN/	/N/	641	2280	87	50	3.75	26	5	2	7
祖	On	/so/	-	581	1127	12	56	3.98	11	9	1	5
俗	On	/zoku/	-	651	576	87	137	4.10	5	9	2	7
賞	On	/shoR/	/R/	577	7931	21	67	4.79	49	15	2	4
害	On	/gai/	-	560	12066	36	62	4.85	8	10	2	4
駅	On	/eki/	-	577	3315	7	44	6.17	5	14	2	3

税	On	/zei/	-	638	22145	20	79	5.25	1	12	2	5
宅	On	/taku/	-	565	10205	36	43	4.19	7	6	2	6
忠	On	/tyuR/	/R/	643	2067	40	53	3.42	13	8	2	6
邸	On	/tei/	-	604	1834	11	16	4.25	22	8	2	7
亭	On	/tei/	-	566	551	5	34	3.83	22	9	2	7
陶	On	/toR/	/R/	735	455	28	29	3.69	37	11	2	7
卓	On	/taku/	-	591	1003	11	30	4.08	7	8	2	7
胴	On	/doR/	/R/	601	187	28	45	5.48	10	10	2	7
念	On	/neN/	/N/	518	9609	40	88	4.04	4	8	2	4
倍	On	/bai/	-	545	4051	87	17	4.40	8	10	2	3
塔	On	/toR/	/R/	634	341	25	32	5.48	37	12	2	7
舶	On	/haku/	-	728	287	6	11	3.92	8	11	2	7
晩	On	/baN/	/N/	594	743	12	52	5.19	4	12	2	6
票	On	/hyoR/	/R/	619	10176	3	19	4.96	7	11	2	4
秒	On	/byoR/	/R/	608	8093	20	5	4.90	5	9	2	3
副	On	/huku/	-	621	11815	24	43	4.04	8	11	2	4
蛮	On	/baN/	/N/	902	74	5	27	3.77	4	12	2	7
幣	On	/hei/	-	776	316	11	42	3.71	11	15	2	7
砲	On	/hoR/	/R/	617	1341	12	50	4.25	22	10	2	7
枚	On	/mai/	-	641	2702	54	17	4.81	4	8	2	6
席	On	/seki/	-	574	10494	11	70	5.69	14	10	2	4
鉄	On	/tetu/	-	540	7040	28	153	6.23	5	13	2	3
脈	On	/myaku/	-	571	702	28	56	5.67	1	10	2	4
盲	On	/moR/	/R/	558	259	15	40	4.31	6	8	2	7
幽	On	/yuR/	/R/	693	95	4	41	3.54	14	9	2	7
郵	On	/yuR/	/R/	688	3436	11	15	3.83	14	11	2	6
肺	On	/hai/	-	548	648	28	21	6.04	10	9	2	6
陸	On	/riku/	-	544	4616	28	64	6.00	1	11	2	4
軸	On	/ziku/	-	649	1353	11	50	5.10	1	12	2	7
猟	On	/ryoR/	/R/	718	188	11	39	5.15	13	11	2	7
缶	On	/kaN/	/N/	582	654	1	7	5.96	44	6	2	7
銀	On	/giN/	/N/	571	13738	28	153	5.98	2	14	2	3
寮	On	/ryoR/	/R/	593	357	36	17	5.33	13	15	2	7
零	On	/rei/	-	599	759	13	18	4.21	12	13	2	7
炉	On	/ro/	-	666	1113	9	36	4.46	3	8	1	7
禅	On	/zeN/	/N/	797	139	12	47	4.15	7	13	2	7
藩	On	/haN/	/N/	681	254	38	40	4.54	20	16	2	7
尿	On	/nyoR/	/R/	615	401	14	31	5.77	3	7	2	7
百	On	/hyaku/	-	595	47834	5	99	5.35	1	6	2	1
銅	On	/doR/	/R/	591	549	28	56	6.02	10	14	2	5
医	On	/i/	-	536	9160	4	59	4.69	21	7	1	3
塾	On	/zuku/	-	555	806	20	12	5.63	2	14	2	7
陽	On	/yoR/	/R/	601	2445	28	60	4.71	20	12	2	3
酪	On	/raku/	-	904	159	10	7	3.79	3	13	2	7
僧	On	/soR/	/R/	718	320	87	118	5.40	33	13	2	7
毒	On	/doku/	-	569	1163	2	76	5.67	3	8	2	4
才	On	/sai/	-	572	644	6	104	3.88	21	3	2	2
列	On	/retu/	-	546	3101	24	76	5.00	4	6	2	3
坪	Kun	/tubo/	-	828	283	25	19	4.92	1	8	2	7
蚊	Kun	/ka/	-	627	81	2	12	6.44	1	10	1	7
滝	Kun	/taki/	-	651	526	103	14	6.23	1	13	2	7
肌	Kun	/hada/	-	610	430	28	49	6.04	1	6	2	7
峠	Kun	/toRge/	/R/	784	107	6	1	5.88	1	9	3	7
又	Kun	/mata/	-	618	247	9	14	3.83	1	2	2	7
塚	Kun	/tuka/	-	712	3787	25	42	4.85	1	12	2	7
芝	Kun	/siba/	-	680	2013	38	38	5.94	1	6	2	7
崎	Kun	/saki/	-	760	10669	6	38	4.15	1	11	2	7
株	Kun	/kabu/	-	687	8839	54	40	5.42	1	10	2	6
娘	Kun	/musume/	-	682	1742	24	33	6.29	1	10	3	7
皿	Kun	/sara/	-	633	486	8	24	5.88	1	5	2	3
芋	Kun	/imo/	-	704	47	38	48	6.25	1	6	2	7
姫	Kun	/hime/	-	643	458	24	56	6.08	1	10	2	7
枠	Kun	/waku/	-	652	2932	54	18	5.00	1	8	2	7
岬	Kun	/misaki/	-	711	156	6	5	6.19	1	8	3	7
棚	Kun	/tana/	-	715	547	54	56	5.73	1	12	2	7
杉	Kun	/sugi/	-	634	2410	54	43	6.06	1	7	2	7
貝	Kun	/kai/	-	584	443	21	90	6.02	1	7	2	1
箱	Kun	/hako/	-	589	1051	21	87	5.92	1	15	2	3
瀬	Kun	/se/	-	701	2530	103	54	4.10	1	19	1	7