Contribution of sublexical information to word meaning: An objective approach using latent semantic analysis and corpus analysis on predicates

Keisuke Inohara (kei.inohara@gmail.com)
1. Department of Informatics, Graduate School of Informatics and Engineering, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Tokyo, 182-8585, JAPAN

Taiji Ueno (taijiueno7@gmail.com)
1. Department of Psychology, Graduate School of Environmental Studies, Nagoya University, Furo-cho, Chikusa-ku, Nagoya City, Aichi 4648601, JAPAN
2. Japan Society for the Promotion of Science

Abstract

Past studies have employed a subjective rating/categorization methodology to investigate whether radicals, an example of sub-lexical visual information in Chinese/kanji, contribute to computation of character/word meaning, with conflicting results. This study took an objective, corpus-based approach for the first time. Specifically, we conducted a Latent Semantic Analysis based on Japanese newspaper text (Experiment 1), and found that radical friends (kanji characters with the same radicals) appeared in more similar linguistic contexts than radical enemies (kanji characters that do not include the same radicals). In addition, we consulted a noun-verb predicate corpus extracted from Japanese web texts (Experiment 2), and showed that nouns including radical friends tended to take more similar predicates than nouns with radical enemies. These findings suggest that characters/words with similar meanings tend to share radicals in kanji, which may explain how children are able to efficiently learn to use the vast number of characters in Chinese/Japanese.

Keywords: semantic radical; latent semantic analysis; predicates; orthography; semantics

Introduction

How word meanings are computed from orthography and phonology (and vice versa) has been a central issue in the psycholinguistic literature. For example, neurocognitive theories differ in whether reading aloud (orthography-phonology mapping) necessarily involves a computation of word meaning (orthography $\rightarrow$ meaning $\rightarrow$ phonology) or not (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Plaut, McClelland, Seidenberg, & Patterson, 1996). To address this issue, one first needs to understand how a word meaning is computed from its written form (Harm & Seidenberg, 2004). Computation of word meaning is also a practical concern in child language acquisition/teaching of both alphabetic (Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001) and non-alphabetic (Hino, Miyamura, & Lupker, 2011) languages. Learning to spell/read a vast number of Chinese characters and Japanese kanji adaptations from Chinese is a demanding problem (Shu, Chen, Anderson, Wu, & Xuan, 2003; Tamaoka & Yamada, 2000). For example, there are 2,136 Japanese kanji characters designated for everyday use. In teaching so many items, a particular emphasis on lexical/semantic associations with orthography/phonology might be effective (NB. To a less extent, English also shares this issue as it does not code each vowel with one specific phoneme).

Although a word meaning is a word-specific type (i.e., whole-word) of knowledge, sub-lexical information also contributes to its computation. Evidence for this has been accumulated in alphabetic languages (Libben, 1998; Marslen-Wilson, Tyler, Waksler, & Older, 1994), and to a greater extent in logographic languages (Hino et al., 2011; Shu et al., 2003). Further insight on this issue has been gleaned by investigating the role of sub-character (visual) information (e.g., radicals) in Chinese and Japanese kanji (Ogawa, 2013; Shu et al., 2003; Tamaoka, 2005). For example, a native Japanese speaker would agree that the kanji characters 洗 (*wash*) and 流 (*flow*) share similar meanings (e.g., water) because these characters share a radical (left part of each character). However, the outcomes of scientific investigations are not consistent with the role of radicals on computing character/word meanings. Specifically, Hino et al. (2011) concluded that orthographic neighbors in kanji compounds (sharing one/two kanji characters, thereby sharing radicals as well) tend to have similar meanings, but the degree of the shared meaning was not different from that of orthographic neighbors in kana (another type of orthography without a radical, which codes phoneme/mora in Japanese), suggesting that the existence of a radical in kanji is not particularly helpful in computing word meanings. Furthermore, other studies suggest there are some exceptional characters whose meanings are different from other characters with the same radicals (Ogawa, 2013; Shu et al., 2003; Tamaoka, 2005). Therefore, it remains unclear whether sub-character information in Chinese/kanji contributes to character/word meaning.

These mixtures of outcomes may stem from the way in which word/character meanings (and semantic similarity) were measured. Specifically, all of these studies employed a subjective rating, such as asking (in a Likert scale) how radical/character neighborhoods are similar in meaning (Hino et al., 2011), or asking or categorizing to what extent each character meaning is consistent with its radical meaning (e.g., in case of the example above, how meaning of 洗 is relevant to water) (Ogawa, 2013; Shu et al., 2003; Tamaoka, 2005). Theses subjective ratings/categorizations could be affected by demand characteristics (Orne, 1962), and by the list composition. Therefore, in this study, we aimed to investigate how radicals predict word meanings by using objective measures of semantic similarity. A closest

example of existing literature is Jin, Carroll, Wu, and MacCarthy (2012), who measured an objective corpus-based semantic similarity between a word pair, and considered the number of shared radicals. However, their target was to explain the subjective semantic similarity measure obtained in human experiments. So, in a strict sense, they did not objectively test whether radical friends were semantically more similar than enemies, which we did in this study.

In Experiment 1, we used latent semantic analysis (Landauer & Dumais, 1997), which has become influential in the last decade. The latent semantic analysis (LSA) assumes that semantic similarity among words/characters can be inferred from the similarity structure of the linguistic context where they appear (i.e., words/characters with similar meanings appear in similar contexts). After establishing that words with the same radicals have similar meanings, i.e., they appear in more similar contexts than those without, we next conducted a finer-grained analysis on their linguistic context similarities in Experiment 2. Specifically, we measured the similarity matrix in the language corpus of predicates. The rationale behind this approach is as follows: Jones (1985) operationalized a word's predicability as the ease with which the word's referent (e.g., ball) "can be described by simple factual statements" (e.g., A ball is *bouncing*). Importantly, this measure has been assumed to be an index of the richness of word meaning (Plaut & Shallice, 1993). A logical consequence is to assume that words sharing similar lists of predicates are semantically similar. Thus, a large corpus of predicates provides another objective test case allowing us to test whether words including kanji characters with the same radicals have more similar lists of predicates than those without. The answer should be yes if radicals contribute to the organization of character/word meanings.

Experiment 1

Radical Friends and Enemies
To begin with, we define two terms (*radical friends & radical enemies*) in order to clarify the experimental procedures. Specifically, a given pair of two kanji characters is termed either *radical friends* or *radical enemies* based on their shared/unshared radicals. For example, the kanji characters in the middle columns of Table 1 share the same radical, shown in the left column [金 (*metal*) for the top row; 木 (*wood*) for the bottom row]. Thus, any pairs of 2 kanji characters within a row are *radical friends*. On the other hand, any pairs of 2 kanji characters from different rows in Table 1 are *radical enemies*.

Table 1. Radicals and kanji exemplars used in Experiment 1

Radical examples (13 radicals)	Kanji exemplars (i.e., radical friends)	Prediction
金 (*metal*)	銀 (*silver*) 銅 (*copper*) 鉄 (*steel*) ...	Radical-friends appear in more similar contexts than radical-enemies.
木 (*wood*)	林 (*forest*) 植 (*planting*) 樹 (*tree*) ...	

We predicted that radical friends appear in more similar linguistic contexts than radical enemies. In order to test this prediction, we calculated *LSA similarities*, indicating similarities between two kanjis in terms of their distributions across contexts.

Materials

Kanji Dataset & Corpus
The materials [radicals & kanji exemplars in each radical cohort (i.e., radical friends)] were selected in the following two steps: (a) First, we focused on 13 highly frequent radicals [top 5% cohort of the radical frequency database (Kondo & Amano, 1999)]. (b) Next, we selected only highly frequent kanji characters in each cohort of the 13 radicals [top 25% cohort of the character frequency database (Amano & Kondo, 2003)]. As a result, 536 kanji characters were selected in total (mean number of the kanji characters

	536 kanji characters		200 numerical values (derived from LSA)						
Target kanji charcter (銀)	銀 (*silver*)	3.32	2.61	1.76	••••	0.99	0.51	••••	
銀's radical-friends: kanji with the same radical as the target kanji character (top row)	鉄 (*steel*)	3.88	2.13	2.08	••••	1.56	0.71	••••	
	釣 (*fishing*)	2.01	1.92	0.75	••••	0.55	0.32	••••	
	...	.	.	.	.	.	.	••••	
	...	.	.	.	.	.	.	••••	
銀's radical-enemies: kanji without the same radical as the target kanji character (top row)	林 (grove)	2.89	3.04	1.58	••••	0.02	0.15	••••	
	植 (*planting*)	0.98	1.24	0.36	••••	0.45	2.58	••••	
	...	.	.	.	.	.	.	••••	
	...	.	.	.	.	.	.	••••	

Figure 1. Illustration of the procedure to compute a mean LSA similarity between 銀 (silver) kanji character and its radical friends (compute cosine vectors between the pairs shown in red arrows and average) and a mean LSA similarity between 銀 (silver) kanji character and its radical enemies (compute cosine vectors between the pairs shown in blue dotted arrows, and average) respectively in Experiment 1.

in each radical cohort was 41.2, *SD* = 21.4, range = 15:82).

In Experiment 1, Japanese newspaper corpora (Mainichi newspaper in 2008) were used (4,156 Japanese unique kanji characters, 38.9 million (token) characters, 1.1 million sentences, and 506.9 thousand paragraphs).

Procedure

LSA (Context Similarities between 2 Kanji Characters)
LSA procedure consists of constructing a kanji-context matrix, performing data reduction, and using similarity between two kanji characters in the reduced matrix as context similarity measure. Thus, estimating the context similarities in LSA involves a compression of a large corpus (see above) into a smaller vector space. For this compression, modellers have to decide two parameters. One parameter is *linguistic context*, referring to language units (e.g., words, sentences, paragraphs, articles, and documents), within which similarities are estimated. For example, suppose we use the sentence as the linguistic context parameter, then a higher similarity in LSA means that two kanji characters tend to appear in the same *sentence* (direct relationship). It also means two kanji characters might not appear simultaneously together within a sentence *but* that each tends to appear in sentences containing similar kanji (indirect relationship). Alternatively, if the paragraph is used as the linguist context parameter, this means two kanji characters tend to appear within the same paragraph. In this study, we conducted 2 LSAs (paragraph and sentence, respectively).

The other parameter is *the dimension (size) of the resultant vector space* after the compression of the corpus. The merit of a smaller dimension size lies in a simpler and less noisy LSA estimation. The smaller dimension, however, means that a larger amount of information is discarded from the original corpus (i.e., trade-off). Our pilot study revealed that a dimension size of 200 was most sensitive.

LSA involves three steps. First, we made a kanji-context matrix (rows = 536 kanji characters; columns = each context unit in the corpus) whose element a_{ij} denoted the number of times the kanji k_i occurred in the j-th context. As explained above, we conducted two LSAs (i.e., two matrices), in which a context unit corresponded to either a sentence or paragraph, respectively. Secondly, the elements of these initial matrices were log-transformed and multiplied by the entropy in order to avoid the effect of enormously frequent appearance (Quesada, 2007). Finally, we decomposed the matrices by singular value decomposition (SVD) and reintegrated them as vector spaces whose kanji k_i ($1 \leqq i \leqq 536$) had a vector of 200 numerical values. The resultant matrix looked like Figure 1.

LSA Similarities of Radical Friends and Enemies
The procedures above allowed us to estimate a LSA similarity between a kanji character and its radical friends/enemies, respectively, by computing a cosine between its row vector and the radical friend's/enemy's row vector, respectively. For example, Figure 1 illustrates the case of 銀 (*silver*) character. In total, 535 cosines can be computed between the target kanji k_i [銀 (silver)] vector and the rest of the character vectors (the other rows). These 535 cosines were split into two: One was the cosines between k_i vector and the vectors of its radical friends [e.g., 鉄 (*steels*), 釣 (*fishing*)], and these were averaged to form a mean LSA similarity over k_i's radical friends; The rest of the cosines [i.e., cosines between k_i vector and the vectors of k_i's radical enemies (e.g., 林 *grove*, 植 *planting*)] were also averaged to form a mean LSA similarity over k_i's radical enemies. This procedure was repeated for each kanji character k_i (each row), resulting in 536 pairs of a mean LSA similarity with radical friends and that with radical enemies, respectively. These mean LSA similarities values were submitted to ANOVA (each one of 536 kanji character as random variables).

	528 kanji characters (Experiment 2)	All the nouns sharing the kanji in the left column (kanji neighborhoods)	Probabilities of 6,612 predicates (values sum to 100% in each row)				
			count	use	build	sing	....
Target kanji charcter (銀)	銀 (*silver*)	銀貨 (*silver coin*) 銀製 (*made by silver*) 白銀 (*platinum*) ...	1.2%	2.7%	0%	0%	....
銀's radical-friends: kanji with the same radical as the target kanji character (top row)	鉄 (*steels*)	鉄道 (*train*) 鉄橋 (*iron bridge*) ...	0.3%	3.0%	0.9%	0%	....
	釣 (*fishing*)	お釣り (*exchange*) 釣針 (*fishing gear*) ...	2.2%	1.4%	0%	0%	....
	...	...	.	.	.	.	....
銀's radical-enemies: kanji without the same radical as the target kanji character (top row)	学 (*learning*)	学校 (*school*) 入学 (*enter a school*) ...	0.02%	1.0%	0%	0%	....
	讃 (*praise*)	讃歌 (*hymn*) 自讃 (*self-respect*) ...	0%	0%	0%	5.4%	....
	...	...	.	.	.	.	....

Figure 2. Illustration of the procedure to compute a mean predicate similarity between 銀 (silver) kanji character and its radical friends (compute cosine vectors between the pairs shown in red arrows and average) and a mean predicate similarity between 銀 (silver) kanji character and its radical enemies (compute cosine vectors between the pairs shown in blue dotted arrows, and average) respectively in Experiment 2.

Results and Discussion

The mean LSA similarities (*SD*) are shown in Table 2. A 2 (linguistic context: paragraph vs. sentence) by 2 (radical: radical friends vs. radical enemies) repeated ANOVA revealed significant main effects of the linguistic context factor ($F(1, 535) = 26.21$, $p < .01$, $\eta_p^2 = 0.047$) and the radicals factor ($F(1, 535) = 17.86$, $p < .01$, $\eta_p^2 = 0.032$). That is, the mean LSA similarities were higher when estimated within a sentence than within a paragraph. More importantly, radical friends had higher similarities than radical enemies, consistent with our predictions. In addition, the interaction between these factors was significant ($F(1, 535) = 5.16$, $p = .024$, $\eta_p^2 = 0.010$). Simple main effects of linguistic context were significant for radical friends ($F(1, 535) = 20.00$, $p < .01$, $\eta_p^2 = 0.036$) and for radical enemies ($F(1, 535) = 21.23$, $p < .01$, $\eta_p^2 = 0.038$), respectively. Simple main effects of the radical factor was also significant in the paragraph-based LSA data ($F(1, 535) = 8.09$, $p < .01$, $\eta_p^2 = 0.015$) and in the sentence-based LSA data ($F(1, 535) = 21.00$, $p < .01$, $\eta_p^2 = 0.038$). The source of the interaction lay in the larger effect of the radical factor (friends vs. enemies) when LSA similarities were estimated within a sentence ($d = 0.198$, Rosenthal, 1991, p.15, equation 2.12) than that within a paragraph ($d = 0.123$).

Table 2. *LSA similarities as a function of linguistic context used in LSA (Paragraph or Sentence) and radical (Friends or Enemies).*

	Paragraph	Sentence
Friends	0.0139 (0.0211)	0.017 (0.0238)
Enemies	0.0122 (0.0149)	0.0139 (0.0177)

notes. LSA = Latent Semantic Analysis; Friends = kanji characters with the same radicals; Enemies = kanji characters that do not share the same radicals.

In summary, radical friends had higher LSA similarities than radical enemies, regardless of the linguistic context size used in LSA, consistent with our hypothesis. That is, radicals provide some useful information for computation of meaning. More interestingly, the effect size of the radical factor (friends > enemies) was stronger in sentences than in paragraph. This indicates that radical friends tend to appear within the same sentence (direct relationship) and/or they tend to co-occur with the same kanji in a sentence (indirect relationship). This fact motivated us to conduct a finer-grained analysis of similarities between radical-friends within a sentence in Experiment 2.

Experiment 2

In Experiment 2, we measured *predicate similarities* between radical friends and compared these with those between radical enemies. In Japanese, most of nouns, which take a predicate, are consisted of two or more kanji/kana characters. Therefore, the following predicate analysis focused on not a single kanji character but nouns that contain a kanji character used in Experiment 1. For example, a kanji character 銀 (*silver*) appears as part of the following nouns and each noun has its list of possible verbs.

銀貨 (*silver coin*) – use, throw, pay, count, etc.
銀行 (*bank*) – go, withdraw, pay, count, etc.

Also, 銀's radical friend 鉄 (*steels*) appears as follows:

鉄道 (*train*) – use, run, count, etc.
鉄砲 (*gun*) – use, count, throw, etc.

In contrast, 銀's radical enemy, 校 (*school*) appears as follows:

高校 (*high school*) – study, enter, etc.
校舎 (*school building*) – study, build, etc.

As one can see above, nouns containing 銀 have similar lists of predicates as nouns containing 鉄 (radical friends) but less similar lists with nouns containing 校 (radical enemies). Experiment 2 investigated these similarity structures objectively by consulting a predicate database (Hayashibe, 2012). The specific prediction was that words including a radical friend kanji character, for instance, 銀貨 (*silver coin*), 銀行 (*bank*) 鉄道 (*train*), 鉄砲 (*gun*) above, would have higher predicate similarities than radical enemies (e.g., 高校 above). As a first step, this study focused on the noun-verb relationship in order to probe the predicate similarity between radical friends and enemies.

Materials

Language Corpus of Predicates

The Noun-Verb predicate corpus (Hayashibe, 2012) is based on about 100 million Japanese web pages (Yata, 2010), and its format is illustrated in Table 3. From this corpus, we extracted the materials for our study in the following steps: First, only the highly frequent noun-verb predicate pairs were extracted (those with a row count of more than 1,000). Next, any noun-verb predicate pairs containing alphabet/number/symbols (e.g., @) were removed. Finally, any noun-verb pairs that did not contain any of the 536 kanji characters used in Experiment 1 were removed. As a result, we obtained 142,457 noun-verb predicate pairs, including 11,652 unique nouns and 6,612 unique verb predicates.

Table 3. *The structure of the language corpus of predicates used in Experiment 2.*

Noun	Predicate (verb)	Frequency
銀貨 (*silver coin*)	count	1,500
銀貨 (*silver coin*)	use	2,000
高校 (*high school*)	use	3,000
...	...	...

Procedure

Creating Probability Matrix of Predicate Occurrences
Like Experiment 1, we made a kanji-predicate matrix consisting of 528 rows (kanji characters) and 6,612 columns (predicates). Eight kanji characters from Experiment 1were not included because they did not appear as part of any nouns in the database we used.

The elements in this kanji-predicate matrix represented the statistical structure of kanji-predicate mappings in Japanese by taking the raw frequency count (see Table 3) into consideration. Specifically, if a noun [e.g., 銀貨 (*silver coin*)] including target kanji k_i [e.g., 銀 (*silver*)] took the predicate p_j (e.g., *use*), we added its token frequency (e.g., 2,000 in Table 3) to element a_{ij}. Then, we divided the resultant a_{ij} value by sum of elements in the k_i vector. In other words, the resultant element a_{ij} denotes the *probability* with which predicate p_j was taken, given that kanji character k_i appeared as part of a noun (see Figure 2 for illustration).

Predicate Similarities of Radical Friends and Enemies
Estimating the predicate similarities between radical friends and radical enemies, respectively, involved almost the same procedure as Experiment 1 (Figure 2). For example, in case of kanji character k_i [e.g., 銀 (*silver*)], cosines between k_i vector (top row in Figure 2) and the rest of the 527 kanji character vectors (other 527 rows) were computed. Then, these 527 cosines were split into two, and averaged to form two means of predicate similarities: One was the mean cosines over k_i's radical friends [e.g., 鉄 (*steels*)]. The other was the mean cosines over k_i's radical enemies [e.g., 学 (*learning*)]. This procedure was repeated for each kanji character k_i (each row), resulting in 528 pairs of a mean predicate similarity with radical friends and with enemies.

Results and Discussion

The mean values (*SD*) of predicate similarities were 0.469 (0.121) between radical friends and 0.450 (0.113) between radical enemies. A paired *t* test with 528 kanji characters as random variables revealed that the mean predicate similarity over radical friends was significantly higher than the mean similarity over radical enemies (t(527) = 10.90, $p < .01$, $d = 0.47$). Note that this difference was still significant even if the probability values in the kanji-predicate matrix were non-linearly transformed (root-squared) to control for outliers ($d = 0.63$). Furthermore, the effect size in Experiment 2 ($d = 0.47$) was higher than that in Experiment 1 ($d = 0.123$ in paragraph; $d = 0.198$ in sentences).

Before discussing the results, it is worth excluding a less interesting interpretation of the current result. Specifically, a high predicate similarity between radical friends could mean either or both of the following:

(1) If Noun X takes Predicate A frequently, then its radical friend (Noun Y) also takes Predicate A frequently.

(2) If Noun X never takes Predicate A, then a radical-friend (Noun Y) never takes Predicate A either.

The result would be less convincing if only the second case was the cause of the higher predicate similarities among radical friends. This interpretation can be rejected, however, by counting the number of (0, 0) pairs when estimating the cosines between a given two vectors, and controlling for its effect on the dependent measure. Following this procedure, an ANCOVA with the number of (0, 0) pairs as a covariate still showed a significant main effect of the radical factor (radical friends vs. enemies), (F(1, 526) = 143.64, $p < .01$, $\eta_p^2 = .215$). This suggests that the higher predicate similarity among radical friends than among radical enemies was not only due to the second case but also due to the existence of common predicates that were taken frequently by radical friends.

General discussion

This is the first objective study to demonstrate these linguistic similarities between radical friends and to provide a statistical structure in radical-predicate mapping. We examined whether radicals, an example of sub-character information in Chinese/kanji, contributed to computation of character/word meaning. We adopted two objective measures of semantic similarities between two kanji characters, whereas previous studies have used subjective measures in estimating semantic similarities [like ratings of similarities on a Likert scale (Hino et al., 2011)]. The outcome of Experiment 1 suggested that kanji characters with the same radical were more similar than those without in terms of the linguistic contexts in which they appear. In Experiment 2, we found that nouns including radical friends had more similar statistical structure of predicates than nouns with radical enemies.

These results indicate that radicals certainly provide useful information for computation of character/word meaning, at least when one generates the target predicate from a noun containing a kanji character. This raises the question of 'how'. One idea would be to assume a probabilistic process. From this viewpoint, the target (semantically congruent) predicate can be computed under multiple probabilistic constraints provided by context, word, character, sublexical information, and so on. Radicals within a noun may serve as one of these probabilistic constraints to compute the target predicate. In fact, the result of Experiment 2 suggests the existence of a graded probabilistic structure between radicals and predicates, rather than a binary structure. Past connectionist models have demonstrated that acquisition/use of knowledge in a quasiregular domain is explained by statistical learning theory (Elman, 1990; Plaut et al., 1996; Seidenberg & McClelland, 1989).

Finally, the role of radicals in computing meaning, particularly in generating the target predicate makes sense given the vast number of characters in Chinese/Japanese kanji compared to alphabetic languages. For instance, the

Japanese newspaper corpora use in Experiment 1 (Mainichi newspaper in 2008) contained 4,156 Japanese different kanji characters while English has only 52 characters ("a" to "z" and "A" to "Z"). Learning this vast amount of kanji characters and their usage (e.g., to generate the target predicate from a kanji-compound word) is a gargantuan task, in which some constraints provided by shared radicals should be useful. In school education, instruction to pay attention to radicals in some kanji characters may have facilitative effects for learning kanji character.

However, one may say the effect sizes were relatively small in both experiments ($d = 0.123$ in paragraph and $d = 0.198$ in sentences in Experiment 1 and $d = 0.47$ in Experiment 2). We would argue this is even adaptive/functional. As we found in Experiment 2, the mapping from a radical within a word to its semantically congruent predicate is an example of a quasiregular domain (Plaut et al., 1996; Seidenberg & McClelland, 1989). If a radical within a word predicts its predicate too strongly, then exceptional/atypical mapping should be hard to learn. Thus, relatively loose constrains from radicals make sense and are functional. Also, an explicit demonstration of such a relatively small effect size and a quasiregular structure (i.e., specifying which radical-predicate pair is exceptional) is useful in teaching. This is because the objective semantic similarity measurement in our study allows teachers to know which radicals benefit from a particular emphasis of its associate meaning whereas which ones do not. Thus, our study should contribute to education by clarifying the boundary condition of a semantically-oriented teaching of radicals.

References

Amano, S., & Kondo, T. (2003). *Nihongo-no Goi-Tokusei (Lexical properties of Japanese) Vol. 7*. Tokyo: Sanseido.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256.

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science, 14*(2), 179-211.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychol Rev, 111*(3), 662-720.

Hayashibe, Y. (2012). *Nihongo kakariuke corpus (Japanese modification corpus)*. nwc2010.verb.th5.bz2 Retrieved from: http://hayashibe.jp/jdc/

Hino, Y., Miyamura, S., & Lupker, S. J. (2011). The nature of orthographic-phonological and orthographic-semantic relationships for Japanese kana and kanji words. *Behavior Research Methods, 43*(4), 1110-1151.

Jin, P., Carroll, J., Wu, Y., & McCarthy, D. (2012). Distributional similarity for chinese: exploiting characters and radicals. *Mathematical Problems in Engineering, 2012*.

Jones, G. V. (1985). Deep dyslexia, imageability, and ease of predication. *Brain and Language, 24*(1), 1-19. Kondo, T., & Amano, S. (1999). *Nihongo-no Goi-Tokusei (Lexical properties of Japanese) Vol. 5*. Tokyo: Sanseido.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Libben, G. (1998). Semantic transparency in the processing of compounds: consequences for representation, processing, and impairment. *Brain and Language, 61*(1), 30-44.

Marslen-Wilson, W., Tyler, L., K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental Lexicon. *Psychological Review, 101*(1), 3-33.

Ogawa, T. (2013). Semantic similarities among Japanese kanji characters sharing same left radicals in the Japanese educational kanji list (Kyoiku kanji list). *Tokai Gakuin Daigaku Kiyou, 6*, 217-223.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776-783.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychol Rev, 103*(1), 56-115.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10*, 377-500.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science, 2*(2 Suppl), 31-74.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Vol. 6): Sage.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523-568.

Shu, H., Chen, X., Anderson, R. C., Wu, N., & Xuan, Y. (2003). Properties of school Chinese: implications for learning to read. *Child Development, 74*(1), 27-47.

Tamaoka, K. (2005). How much do 'water' and 'string' radicals contribute to kanji meanings? *Bulletin of the International Student Center, Hiroshima University, 15*, 11-24.

Tamaoka, K., & Yamada, H. (2000). The effects of stroke order and radicals on the knowledge of Japanese kanji orthography, phonology and semantics. *Psychologia, 43*(3), 199-210.

Yata, S. (2010). *Japanese web corpus 2010*. Retrieved from: http://s-yata.jp/corpus/nwc2010/