Discovering Phonesthemic Clusters in Readings of Kanji Characters toward Exploring Phonestheme in Japanese

Akira Yoshida[1*] **Chihaya Matsuhira**[1] **Hirotaka Kato**[1] **Takatsugu Hirayama**[2,1]
Takahiro Komamizu[1] **Ichiro Ide**[1]
[1]Nagoya University [2]University of Human Environments
{yoshidaa,matsuhirac}@cs.is.i.nagoya-u.ac.jp
hirotaka.kato@mirai.nagoya-u.ac.jp
t-hirayama@uhe.ac.jp, taka-coma@acm.org, ide@i.nagoya-u.ac.jp

Abstract

Phonestheme is a particular sequence of speech sounds that conveys a certain meaning. Most of the studies on identifying phonestheme have focused mainly on European languages, but few studies have focused on Japanese. However, since phonestheme plays an important role in language acquisition and brand naming, discovering a concept similar to phonestheme also in Japanese can be beneficial in Japanese education and commerce. In this paper, we hypothesize the existence of a concept named "phonesthemic clusters" in Japanese Kanji characters and propose a method to identify them by focusing on the consonants of the readings of Kanji characters. We apply the proposed method to 2,136 common Kanji characters and show that it successfully extracts 100 and 33 phonesthemic clusters in the first and second consonants of reading, respectively. Also, the proposed method automatically labels the extracted phonesthemic clusters, successfully assigning semantic labels to more than 80% of them. These results suggest the existence of the systematic correspondence between consonants and meanings in the reading of Japanese Kanji characters.

1 Introduction

Phonestheme is a particular sequence of speech sounds that conveys a certain meaning, which was first proposed by a linguist Firth (1930). For example, English words beginning with a phonestheme "gl" include many words related to light, such as "gleam" and "glitter", and thus the phonestheme "gl" is considered to have a meaning related to light (Berge, 2004). Such phonesthemes are known to have an important role in language acquisition and brand naming, which implies their usefulness in education and commerce (McCune, 2011; Parault and Schwanenflugel, 2006). Phonesthemes have been studied mainly in European languages such as English and Swedish (Åsa Abelin, 1999) but few studies have focused on Japanese. One of the reasons for this is that Japanese has a phonological limitation of not having a sequence of consonants, whereas phonesthemes are generally composed of two or three consonants. Among the few studies, Hamano (1998) analyzed the correspondence between phonemes and meanings in Japanese ideophone. However, because this analysis was limited to ideophone, their meanings were not as specific as those of phonesthemes in English. In addition, since Hamano's evaluation was conducted manually by the author, the obtained results were highly labor-intensive.

Therefore, through data-driven approach, this paper attempts to automatically discover a concept similar to phonesthemes also in Japanese which we call "phonesthemic clusters". We focus on the Sino-Japanese readings of Japanese Kanji characters[1] assuming that they express more specific meanings than ideophone. Specifically, this paper targets consonants in Japanese readings of Kanji characters. The reading always has either form of CV, CVC, or CVCV, where C and V stand for a Japanese consonant and a Japanese vowel, respectively. We define a "phonesthemic cluster" as a cluster of Kanji characters that share specific semantics and show high proportion of a specific consonant either in positions of the first or second consonant as a stepping stone to identifying phonestheme in Japanese. Our motivation comes from the fact that there exists some examples that Kanji

*The author is currently affiliated with DeNA Co., Ltd.

[1]In general, Japanese Kanji characters have two types of reading: the Japanese reading (*kun-yomi*) and Sino-Japanese reading (*on-yomi*) which is derived from old Chinese. In this paper, we focus on the latter type of reading.

characters having a specific meaning tend to share a specific consonant in their reading. For example, Kanji characters related to colors often have 'k' in the second consonant such as 白 [haku] (white), 黒 [koku] (black), 赤 [seki] (red), and 緑 [ryoku] (green). Another example is that those having an aggressive sense share 't' in the second consonant such as 殺 [satu] (kill), 滅 [metu] (destroy), 罰 [batu] (punish), and 切 [setu] (cut). The reason for focusing on the readings of Kanji characters is that it is easier to analyze than other readings or words in general because it is always mono- or bimoraic, more systematized, and less influenced by dialects.

The contribution of this paper can be summarized as follows:

- We propose a method to discover phonesthemic clusters in Japanese, especially in readings of Kanji characters, and a method to automatically label these clusters with appropriate meanings in a data-driven approach,

- We apply the proposed method to 2,136 common Kanji characters which results in finding 133 phonesthemic clusters of Kanji sharing both a specific consonant and a meaning,

- We compare the identified phonesthemic clusters with phonesthemes attested in other languages to confirm the similarity of the relationship between a specific consonant and a meaning across languages.

2 Related Work

2.1 Automatic identification of phonestheme

Otis and Sagi (2008) proposed a method for automatically identifying English phonesthemes. Many of the earlier studies manually evaluated the relationship between meanings and pronunciations, which led to analysis only on a small scale. They enabled automatic identification of the existence of many phonesthemes by converting English words into semantic embeddings and testing the average distance between words sharing a phonestheme and randomly selected words.

However, most of the studies of phonestheme aimed at identifying the existence of known phonesthemes, and the identification of novel phonesthemes is underexplored. Liu et al. (2018) attempted to automatically discover both known and novel phonesthemes in English using a linear regression and sparse regularization. They evaluated the identified phonesthemes by recruiting native English speakers and asking them to judge how well each phonestheme fits its meaning, which resulted in the conclusion that phonesthemes could be extracted automatically.

In the method proposed by Otis and Sagi (2008), they first grouped English words by known phonesthemes, and then performed a t-test to the distance between the groups and a group of randomly selected words. Since there are no known phonesthemes in Japanese Kanji characters, it is difficult to directly adopt this method to them.

2.2 Automatic labeling of meaning of phomestheme

Automatic identification of the semantic meaning of each phonestheme was one of the issues raised by Otis and Sagi (2008). Abramova et al. (2013) used the English concept dictionary WordNet (Fellbaum, 1998) to automatically identify the meaning of phonesthemes. Specifically, first, for every noun in an English word cluster that shares a specific phonestheme, hypernyms in WordNet are extracted and used as candidate labels for the cluster. Next, for each candidate label, an affinity score is calculated between the label and each word in the cluster. This score is inversely proportional to the distance between the two words in the WordNet hierarchy. The label with a higher score is judged to represent the meaning of the entire cluster and adopted as a semantic label.

In this paper, we extend their method to identify the meanings of phonesthemic clusters in Japanese.

3 Discovering Process of Phonesthemic Clusters

In this section, we introduce a method for discovering phonesthemic clusters using semantic embeddings of Japanese Kanji characters and a population proportion test. In particular, based on Otis and Sagi's method, we first attempt clustering Kanji characters on the semantic space to obtain *semantic clusters* and then analyzing the bias of consonants by a population proportion test to obtain phonesthemic clusters. Figure 1 shows the procedure of the proposed method.

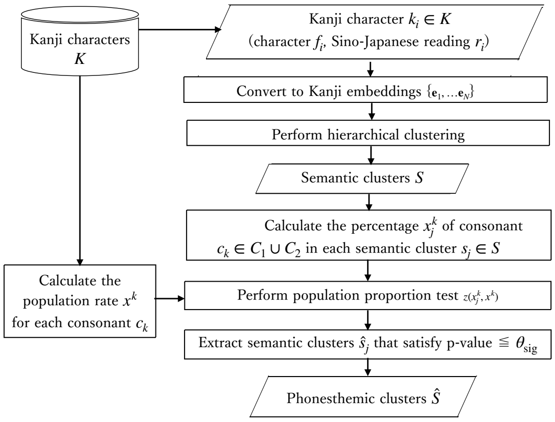

Figure 1: Proposed procedure of discovering phonesthemic clusters.

3.1 Semantic Kanji embedding

Each element k_i of the set of Kanji characters $K = \{k_1, k_2, \ldots, k_N\}$ is transformed into its Kanji embedding $\mathbf{e}_i$ using a pretrained language model. In this paper, we use Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) pretrained on Japanese Wikipedia articles (Inui Lab, 2022).

3.2 Extraction of semantic clusters

As introduced in Section 1, it is known that English words that share a phonestheme express similar meanings. Based on the assumption that this is also the case in Japanese phonesthemic clusters, we apply a hierarchical clustering method (Joe. H. Ward Jr., 1963) to the Kanji embeddings $\mathbf{e}_1, \ldots \mathbf{e}_N$. The set of all Kanji clusters that appear in the process is denoted as S'. Next, for each element s'_j of S', those that satisfy $|s'_j| \geq \theta_{\mathrm{size}}$ are extracted as the semantic cluster s_j and are considered as candidates for phonesthemic clusters. A set of semantic clusters is denoted as S, such that, $S = \{s'_j | s' \in S', |s'| \geq \theta_{\mathrm{size}}\}$. Here, θ_{size} should be maximized to ensure that each cluster contains a sufficient number of elements for a statistical test.

Note that clusters appearing throughout the process of agglomerating the clusters are treated as independent clusters. For example, if {白[haku] (white), 黒[koku] (black), 赤[seki] (red)} and {白[haku] (white), 黒[koku] (black), 赤[seki] (red), 緑[ryoku] (green), 青[sei] (blue), 黄[kou / ou] (yellow)} appear in the clustering process, these two clusters are treated as separate semantic clusters. This is because Kanji characters within each cluster should be semantically related, regardless of the size of a cluster.

3.3 Extraction of phonesthemic clusters

To discover phonesthemic clusters, a population proportion test for the bias of consonants is performed in the following steps. First, given C_1 and C_2 are sets of the possible first and second consonants in the reading, respectively, we first calculate the population proportion x^k of a consonant $c_k \in C_1 \cup C_2$ in the set of whole Kanji characters. Second, the proportion x_j^k of each consonant c_k is calculated within each semantic cluster $s_j \in S$. Third, for the proportion x_j^k of each consonant c_k in each cluster s_j, a population proportion test is performed. The test statistic z is calculated by the following formula:

$$z(x_j^k, x^k) = \frac{x_j^k - x^k}{\sqrt{\frac{x^k(1-x^k)}{|s_j|}}}. \qquad (1)$$

Assuming that this z follows the standard normal distribution $\mathcal{N}(0, 1)$ based on the central limit theorem, we refer to the standard normal distribution table to calculate a p-value.

Finally, for any consonant c_k, the semantic cluster s_j whose p-value satisfies the significance level θ_{sig} is judged to be biased toward the consonant and is identified as a phonesthemic cluster $\hat{s}_j$. This $\hat{s}_j$ can be formulated as follows:

$$\hat{s}_j = \{s_j \in S \mid \exists c_k \in C_1 \cup C_2 : \ z(x_j^k, x^k) \leq \theta_{\mathrm{sig}}\}. \quad (2)$$

A one-tailed test is performed since we want to find semantic clusters of Kanji characters that contain a significantly large number of specific consonants in specific positions. For each s_j, we define a representative consonant $c_r \in C_n$ which meets the condition $z(x_j^r, x^r) = \min_r z(x_j^r, x^r)$. If two or more semantic clusters in a subset-superset relationship satisfy the significance level, the one with the smallest number of elements is extracted as the phonesthemic cluster. For example, if all of the three semantic clusters {白[haku] (white), 黒[koku] (black), 赤[seki] (red)}, {白[haku], 黒[koku], 赤[seki], 青[sei] (blue)} and {白[haku], 黒[koku], 赤[seki], 青[sei], 緑[ryoku] (green), 黄[kou / ou] (yellow)} satisfy the significance level, only {白[haku], 黒[koku], 赤[seki]} is treated as a phonesthemic cluster. This is due to the assumption that a cluster with fewer elements has stronger semantic relationship. The set of phonesthemic clusters extracted from the above method is denoted as $\hat{S} \subseteq S$.

4 Automatic Labeling of Phonesthemic Clusters using WordNet

This section introduces a labeling method based on Abramova et al. (2013)'s one. Since their method was originally proposed for English words, we extend the method by using the Japanese WordNet (Bond et al., 2009) and make it possible to accept a set of Japanese Kanji characters as input instead of a set of English words. Specifically, input of a Kanji character is converted into corresponding English words, such as a Kanji character "犬" being converted into both "dog" and "spy".

4.1 Finding corresponding WordNet nodes for each Kanji

For each Kanji character $k_i \in \hat{s}_j$, we search for corresponding nodes in WordNet (Fellbaum, 1998) using Natural Languate ToolKit (NLTK) (Bird et al., 2009). WordNet is a thesaurus defining hierarchical relationships between English words. Specifically, when a Kanji character k_i is input, NLTK retrieves the corresponding English word node in the WordNet hierarchy. For example, from the Kanji character "白 (white)", the nodes for English words "white" and "whiteness" are retrieved.

4.2 Calculation of affinity score

Next, for each node, the set of hypernyms in WordNet is retrieved. For each Kanji character k_i, we define $H(k_i) = \{k_i\} \cup h(k_i)$ as the set of k_i and its hypernyms $h(k_i)$. Here, we exclude nodes of overly abstract nouns that are distant from the root node by less than a threshold θ_{dist} in the depth of the WordNet hierarchy. Next, for each phonesthemic cluster $\hat{s}_j$, the union of all $H(k_i)$ is calculated as $L_{\hat{s}_j} = \bigcup_{k_i \in \hat{s}_j} H(k_i)$. We regard all elements $l_{\hat{s}_j} \in L_{\hat{s}_j}$ as candidates for semantic labels, and for each of them, the affinity score $A(l_m, \hat{s}_j)$ between a semantic label l_m and a phonesthemic cluster $\hat{s}_j$ is calculated by the following formula:

$$A(l_m, \hat{s}_j) = \sum_{k_i \in \hat{s}_j} \alpha(k_i, l_m), \quad (3)$$

Table 1: Population proportion of each first consonant in common Kanji data.

k	g	s	z	t
0.201	0.052	0.217	0.063	0.108
d	**n**	**h**	**b**	**m**
0.025	0.017	0.088	0.046	0.030
y	**r**	**w**	**ϕ**	
0.031	0.053	0.002	0.068	

Table 2: Population proportion of each second consonant in common Kanji data.

k	t	N	ϕ
0.108	0.056	0.205	0.631

where $\alpha(k_i, l_m)$ is given by

$$\alpha(k_i, l_m) = \begin{cases} \frac{1}{\text{dist}(k_i, l_m)^2} & \text{if } l_m \in H(k_i) \\ -g & \text{otherwise} \end{cases}, \quad (4)$$

where g is a constant representing a penalty, and $\text{dist}(k_i, l_m)$ represents the length of the shortest path from k_i to l_m in the WordNet hierarchy and returns 1 if k_i and l_m are identical. Among the candidate semantic labels, those with a positive affinity score and a coverage $p_{\text{cover}}(l_m)$ greater than a threshold θ_{cover} are adopted as the semantic labels for the phonesthemic cluster. Here, $p_{\text{cover}}(l_m)$ represents the percentage of Kanji characters in a phonesthemic cluster that have l_m as their hypernym, which is calculated using the following formula:

$$p_{\text{cover}}(l_m) = \frac{|\{k_i \mid l_m \in H(k_i), k_i \in \hat{s}_j\}|}{|\{k_i \in \hat{s}_j\}|}. \quad (5)$$

5 Experiment

In this section, we report the results of the experiment in which we applied the proposed method to actual Japanese Kanji data.

5.1 Experimental settings

Kanji data First, we prepared 2,136 common Kanji characters paired with their readings from the Joyo Kanji Table[2] defined by the Japanese government for use in daily life. The consonants of

[2] https://www.bunka.go.jp/kokugo_nihongo/sisaku/joho/joho/kijun/naikaku/pdf/joyokanjihyo_20101130.pdf (Accessed: 2022-12-18)

Table 3: Examples of phonesthemic clusters discovered by the proposed method. Z-scores represent the test statistic for the population proportion test, and for each phonesthemic cluster, Kanji characters including the representative consonants are presented in bold.

Position	Phonesthemic Cluster	Representative Consonant	Proportion	z-score
First Consonant	**価**(ka) **果**(ka) **均**(kiN) **献**(keN / koN) **貢**(kou / ku) **勲**(kuN) **功**(kou / ku) 値(ti) 績(seki)	k	7/ 9	4.311
	講(kou) **教**(kyou) **校**(kou) **究**(kyuu) **研**(keN) 学(gaku) 祉(si) 術(zyutu) 療(ryou)	k	5/ 9	2.649
	清(sei) **静**(sei) **聖**(sei) **神**(siN) **仙**(seN) 浄(jou) 鎮(tiN) 宮(kyuu) 天(teN)	s	5/ 9	2.460
	蚕(saN) **絹**(keN) **繭**(keN) **繊**(seN) **藍**(raN) **綿**(meN) 紡(bou) 糸(si) 桑(sou) 麻(ma)	m	2/10	2.153
Second Consonant	**結**(ketu) **接**(setu) **雑**(zatu / zou) **密**(mitu) 携(kei) 連(reN) 関(kaN) 係(kei) 絡(raku)	t	4/ 9	5.083
	黒(koku) **白**(haku) **赤**(seki) **緑**(ryoku) 青(sei) 紫(si) 紅(kou) 黄(kou)	k	4/ 8	3.601
	蚕(saN) **絹**(keN) **繭**(keN) **繊**(seN) **藍**(raN) **綿**(meN) 紡(bou) 糸(si) 桑(sou) 麻(ma)	N	6/10	3.111
	進(siN) **信**(siN) **伝**(deN) **運**(uN) **展**(teN) 流(ryuu) 交(kou) 情(jouu / sei) 集(syuu) 達(tatu)	N	5/10	2.315

readings were determined based on the "Roman spelling system" (*Kunrei-shiki*). Here, Kanji characters whose Sino-Japanese reading is not listed on the table are treated as having no consonants, which is represented as "ϕ". Moraic nasal consonants represented by "ン(N)" are indicated using "N", whereas other dental nasal consonants in "ナ(na), ニ(ni), ヌ(nu), ネ(ne), ノ(no)" are indicated using "n". Out of the 2,136 characters, 2,062 have at least one, and 152 have at least two readings. For the 74 characters that have no reading, we treated them to have a reading "ϕ". We omitted 1 Kanji character that could not be converted to its Kanji embedding because of the absence of the Kanji character in the model's vocabulary. In consequence, 2,135 Kanji characters were used in this expeirment.

Hyperparameters For the hyperparameters, we empirically set $\theta_{\text{size}} = 7$, $\theta_{\text{sig}} = 0.025$, $g = -0.10$, $\theta_{\text{dist}} = 3$, and $\theta_{\text{cover}} = 0.2$.

5.2 Result of discovered phonesthemic clusters

Hierarchical clustering on the 2,135 Kanji characters resulted in a total of 2,134 semantic clusters. Then, the population proportion test of the first and second consonants was performed targeting between the semantic clusters and the population of the 2,135 common Kanji characters. Tables 1 and 2 show the population proportions of the first and second consonants, respectively. As a result, out of the 2,134 semantic clusters, 100 for the first consonant and 33 for the second consonant were extracted as phonesthemic clusters, respectively. Table 3 shows examples of the discovered phonesthemic clusters (Full lists are provided in the appendix). The results suggest the existence of correspondences between consonants and meanings in reading, which is in line with our expectation.

Comparing the discovered phonesthemic clusters between the first and second consonants, 18 clusters were common and 4 clusters were inclusive. For example, for the cluster {蚕(saN) (silkworm), 絹(keN) (silk), 繭(keN) (cocoon), 繊(seN) (fiber), 藍(raN) (indigo), 綿(meN) (cotton), 紡(bou) (spinning), 糸(si) (yarn), 桑(sou) (mulberry), 麻(ma) (hemp)}, both the first consonant "m" and the second consonant "N" satisfied the significance level. However, it is unlikely that the consonant pair (m, N) plays a phonesthemic role in this cluster because there is only one Kanji character ("綿(meN) (cotton)") in the cluster that has the first consonant "m" and the second consonant "N". Also, since each first consonant belonged to an average of 7.7 phonesthemic clusters and each second consonant belonged to an average of 11.0 phonesthemic clusters, the meaning of each consonant could not be determined uniquely. For example, there were 15 clusters that shared the representative second consonant "N", such as the one with "蚕(saN) (silkworm)" and another with "進(siN) (advance)" listed in Table 3. In contrast to the definition of phonesthemes to evoke specific meanings, too many meanings associated with each consonant may lead to a dilution of the relationship between the consonant and its

Table 4: Examples of semantic labels automatically assigned to the discovered phonesthemic clusters.

Phonesthemic Cluster	Conceptual Dictionary-based Method		
	Semantic Label	Coverage Rate	Affinity Score
黒 (black) 白 (white) 青 (blue) 赤 (red) 紫 (purple) 緑 (green) 紅 (red) 黄 (yellow)	spectral color	5/ 8	4.70
	color	7/ 8	1.65
	piece	2/ 8	1.40
割 (cut) 擦 (rub) 裂 (crack) 張 (chang) 貼 (paste) 塗 (coating) 削 (peel) 剥 (peel)	—	—	—
滴 (drip) 晶 (crystal) 泡 (bubble) 豆 (bean) 穀 (drain) 麦 (wheat) 粉 (flour) 菓 (sweets) 粒 (drains)	grain	3/ 9	2.40
	sphere	3/ 9	1.51
	food product	4/ 9	1.25
短 (short) 長 (long) 少 (small) 中 (medium) 小 (small) 大 (large) 半 (half) 高 (high) 低 (low)	size	2/ 9	1.30
	concept	2/ 9	0.55
	—	—	—
置 (place) 立 (stand) 持 (have) 存 (exist) 産 (birth) 生 (birth) 出 (produce) 入 (enter) 成 (produce) 行 (go) 発 (emerge)	act	3/11	1.31

meaning. Since we try to discover phonesthemic clusters as a stepping stone to identifying phonesthemes in Japanese, further analysis of the phonesthemic consonants in Kanji readings, prioritizing the meanings associated with each consonant is required in future work.

5.3 Result of automatic labeling

Applying our automatic labeling to the discovered phonesthemic clusters resulted in assigning one or more semantic labels to 84 out of 100 phonesthemic clusters represented by the first consonants and 30 out of 33 phonesthemic clusters represented by the second consonants.

Table 4 shows examples of the labels automatically assigned to each phonesthemic cluster. For example, the phonesthemic cluster {価[ka] (value), 値[ti] (value), 果[ka] (fruit), 均[kiN] (equal), 献[keN / koN] (donation), 貢[kou / ku] (tribute), 勲[kuN] (merit), 功[kou / ku] (merit), 績[seki] (achievement)} was labeled with the meanings of "value" and "exploit", and {携[kei] (involve), 連[reN] (relation), 関[kaN] (link), 係[kei] (link), 絡[raku] (connection), 結[ketu] (connection, 接[setu] (connection), 雑[zatu / zou] (miscellaneous), 密[mitu] (density)} was labeled with the meanings of "connection" and "relate". This indicates that phonesthemic clusters were identified in accordance with the meanings of the Kanji characters.

5.4 Cross-lingual comparison

Phonestheme is generally considered to be a language-specific phenomenon (Dwight L. Bolinger, 1950), but some phonesthemes have been reported to be common across languages such as English and Swedish (Åsa Abelin, 1999). In this section, we compare the phonesthemic clusters identified using the proposed method with the phonesthemes reported in various languages, and discuss the similarities and differences among them.

5.4.1 Comparison with phonesthemes in English

We first compare our phonesthemic clusters with the English phonesthemes reported by Magnus (2001). Magnus grouped English monosyllabic words according to whether or not they contained each consonant, and categorized the meanings of the words in each group. Then, for each meaning category, the percentage of words in the group that represent the meaning was calculated. We use these data of consonants attatched with the percentage of each meaning. Next, we classify each of the phonesthemic clusters into groups according to their representative consonant. For example, the phonesthemic cluster {講[kou] (lecture), 教[kyou] (teaching), 校[kou] (school), 究[kyuu] (research), 研[keN] (study),

Table 5: Examples of phonesthemic clusters with relatively high affinity scores to the meaning of consonants in English. Original expressions of consonants in Magnus (2001) are displayed next to the representative consonants with brackets.

Phonesthemic Cluster	Representative Consonant	Meaning in English	Meaning Percentage	Affinity Score
閣(kaku) **官**(kaN) **監**(kaN) **揮**(ki) 将(syou) 臣(siN / ziN) 督(toku) 令(rei)	k（/k/）	control	8.0%	1.40
始(si) **初**(syo) **終**(syuu) **先**(seN) 次(zi) 継(kei) 続(zoku) 完(kaN) 末(matu)	s（/s/）	start	7.9%	1.30
岸(gaN) **岳**(gaku) 頂(tyou) 畔(haN) 麓(roku) 峡(kyou) 渓(kei) 峠 岬(kou) 峰(hou)	g（g/）	valleys	13.0%	0.45
張(tyou) **貼**(teN) **塗**(to) 割(katu) 擦(satu) 裂(retu) 削(saku) 剥(haku)	t（/t/）	touch	25.7%	0.30
連(reN) **絡**(raku) 携(kei) 関(kaN) 係(kei) 結(ketu) 接(setu) 雑(zatu) 密(mitu)	r（/r/）	connections	14.3%	0.20
言(geN・goN) **語**(go) **号**(gou) 音(ne) 声(sei) 字(zi) 呼(ko) 称(syou) 名(mei) 番(baN)	g（/g/）	voice	3.7%	0.10

学(gaku) (study), 祉(si) (welfare), 術(zyutu) (art), 療(ryou) (therapy)} and {価(ka) (value), 果(ka) (fruit), 均(kiN) (equal), 献(keN / koN) (donation), 貢(kou / ku) (tribute), 勲(kuN) (merit), 功(kou / ku) (merit), 値(ti) (value), 績(seki) (achievement)} are classified into the group of the consonant "k". Next, for each consonant group, we compare the meanings of each consonant in English (Magnus, 2001) and the labels assigned to the phonesthemic clusters obtained by the proposed method.

The comparison is made using Equation 3, which calculates an affinity score between a semantic label and a phonesthemic cluster, as used in the labeling method in Section 4. Specifically, we regard the meaning of a consonant in English as a candidate semantic label l_m and compute the affinity score $A(l_m, \hat{s}_j)$. Unlike in Section 4, we do not set the requirement of positive values for the affinity scores, nor do we set a threshold for the coverage ratio or the shortest distance from the root node. A higher affnity score implies a higher semantic relation between the meaning of a phonestheme in English and that of each Kanji character in a phonesthemic cluster since it indicates a closer distance in the WordNet hierarchy between the candidate semantic label and each Kanji in the cluster.

Although most of the affinity scores between the candidate semantic labels and the phonesthemic clusters were highly negative, some of the candidate semantic labels showed relatively high affinity scores with the phonesthemic clusters that contain specific consonants significantly more, which are shown in Table 5. The mapping between English consonants and consonants of readings of Kanji characters were determined according to Kindaichi (2010). For example, in Magnus (2001)'s work, 8.0% of the English words containing the consonant /k/ were considered to express the meaning of "control," and the affinity score between "control" and {閣(kaku) (cabinet). 官(kaN) (government), 監(kaN) (supervisor), 揮(ki) (command), 将(syou) (general), 臣(siN / ziN) (minister), 督(toku) (governer), 令(rei) (order)}, the phonesthemic cluster containing significantly more "k", was 1.40, which is a relatively high value. This suggests that the consonant /k/ in English and the consonant "k" in the reading of Kanji characters convey a similar meaning.

On the other hand, the meaning percentage in some clusters is less than 10%, which raises a question whether each representative consonant truly represents its corresponding meaning. To avoid using weaker sound-meaning relationships, the necessity of filtering English meanings used for this comparison remains to be considered.

5.4.2 Comparison with phonesthemes in various languages

Next, we also compared phonesthemes attested in various languages (Plato, 1999; Hamano, 1998) with the phonesthemic clusters in the same procedure. Some of the results are shown in Table 6.

The correspondence between Greek letters and the alphabet was determined based on the translation by Reeve (Plato, 1998). There were one first consonant and eight second consonants that showed relatively high affinity scores between the meanings of the Greek phonesthemes (Plato, 1999) and the phonesthemic clusters. For example, Plato (1999) assumed that the consonant "τ" in Greek represents the meaning of "binding," and its affinity score to the phonesthemic cluster {張(tyou)

Table 6: Examples of phonesthemic clusters and consonant meanings in various languages with relatively high affinity scores.

Phonesthemic Cluster	Other Language	Representative Consonant	Meanings in Other Language	Affinity Scores
張(tyou) **貼**(teN) **塗**(to) 割(katu) 擦(satu) 裂(retu) 削(saku) 剥(haku)	Greek (Plato, 1999)	t（τ）	binding	1.40
厳(goN) **激**(geki) **極**(kyoku・goku) 緩(kaN) 快(kai) 急(kyuu) 著(tyo) 微(bi)	Japanese ideophone (Hamano, 1998)	g	hard	0.29
厚(kou) **強**(kyou) **軽**(kei) **堅**(keN) **固**(ko) 深(siN) 浅(seN) 薄(haku) 余(yo) 弱(zyaku) 重(zyuu)	Japanese ideophone (Hamano, 1998)	k	hard, heavy	0.00
追(tui) **超**(tyou) **突**(totu) 引(iN) 巻(kaN) 押(ou) 越(etu) 衝(syou)	Japanese ideophone (Hamano, 1998)	t	movement	−0.45
鮮(seN) **摂**(setu) **触**(syoku) **染**(seN) **掃**(sou) **織**(syoku) 濃(nou) 暗(aN) 汚(o) 汽(ki) 淡(taN) 喫(kitu) 吸(kyuu)	Japanese ideophone (Hamano, 1998)	s	gliding movement	−1.30
鮮(seN) **摂**(setu) **触**(syoku) **染**(seN) **掃**(sou) **織**(syoku) 濃(nou) 暗(aN) 汚(o) 汽(ki) 淡(taN) 喫(kitu) 吸(kyuu)	Greek (Plato, 1999)	s（σ）	wind	−1.30

(stretch), 貼(teN) (paste), 塗(to) (paint), 割(katu) (split), 擦(satu) (rub), 裂(retu) (split), 削(saku) (shave), 剥(haku) (peel)} which contains significantly more first consonant "t", was 1.40, a relatively high value.

The comparison with Hamano (1998) revealed 23 Japanese consonants that have relatively high affinity scores. For example, Hamano found that the first consonant "g" in Japanese ideophone represents the meaning "hard," and the affinity score between "hard" and the phonesthemic cluster that contains significantly more "g" {緩(kaN) (slow), 快(kai) (fast), 急(kyuu) (rapid), 厳(geN / goN) (severe), 激(geki) (intense), 著(tyo) (striking), 極(kyoku / goku) (extreme), 微(bi) (slight)} showed a value of 0.29 which is a relatively high score because about 85% of the phonesthemic clusters did not have positive values.

Among the meanings of the phonesthemes in German (W. von Humboldt, 1836), none of them showed relatively high affinity scores with any phonesthemic cluster.

In the above comparison between various languages, it was confirmed that "g", "k", "s", and "t" were the most common consonants that showed high affinity scores. The common feature of "g", "k", and "t" is that they are plosive sounds. As they tended to show high affinity scores, it could be implied that plosive sounds have stronger sound-meaning relationship than the other types of language sounds. However, the number of data on the meanings of the phonesthemes in Japanese ideophone, German, and Greek are extremely small compared to that in English. Therefore, increasing the number of phonesthemes and their meanings in these languages would be our future work to obtain more reliable results. Also, considering the historical fact that the readings are derived from old Chinese, it would be interesting to compare our findings and the consonant-meaning correspondences attested in modern or old Chinese.

6 Conclusion

In this paper, we hypothesized the existence of a concept similar to phonestheme named "phonesthemic cluster" in Japanese Kanji characters. We proposed a method to discover phonesthemic clusters by focusing on readings of Kanji characters and a method to automatically label their meanings through a data-driven approach. For identifying phonesthemic clusters, we first converted Kanji characters into Kanji embeddings, and then extracted semantic clusters of Kanji characters by hierarchical clustering. Then, a population proportion test was performed on the first and second consonants of the Japansese reading of the Kanji characters included in each cluster, where clusters containing a specific consonant significantly more were identified as phonesthemic clusters. We applied the proposed method to 2,136 common Kanji characters and discovered 100 and 33 phonesthemic clusters for the first and second consonants, respectively. The automatic labeling successfully assigned semantic labels to more than 80% of the extracted phonesthemic clusters. These results suggest the existence of the correspondence between certain consonants and meanings in the readings of Kanji characters. The results also suggest that the first and second consonants of the readings can function as phonesthemes in Japanese.

This paper focused only on the consonants in the readings of Kanji characters to discover phonesthemic clusters. However, by carefully selecting words to be analyzed, we believe that the proposed method can be applied to other types of Kanji readings and even more general Japanese vocabulary. Expanding the scope of our analysis would enhance the understanding of phonestheme in Japanese.

Acknowledgements

Part of this work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (22H03612).

References

Ekaterina Abramova, Raquel Fernández, and Federico Sangati. 2013. Automatic labeling of phonesthemic senses. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, pages 1696–1701, Berlin, Germany.

Benjamin K. Berge. 2004. The psychological reality of phonaesthemes. *Language*, 80(2):290–311.

Steven Bird, Edward Loper, and Ewan Klein. 2009. *Natural Language Processing with Python*. O'Reilly Media, Inc., Sebastopol, CA, USA.

Francis Bond, Hitoshi Isahara, Kiyotaka Uchimoto, Takayuki Kuribayashi, and Kyoko Kanzaki. 2009. Extending the Japanese WordNet. In *Proceedings of the 15th Annual Meeting of the Association for Natural Language Processing*, pages 80–83, Tottori, Japan.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. *BERT: Pre-training of deep bidirectional transformers for language understanding*. *In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol.1, pages 4171–4186, Minneapolis, MN, USA.

Dwight L. Bolinger. 1950. Rime, assonance and morpheme analysis. *Word*, 6(2):117–136.

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA, USA.

John Firth. 1930. *Speech*. Oxford University Press, NewYork, NY, USA.

Shoko Hamano. 1998. *The Sound-Symbolic System of Japanese*. CSLI Publications, Springfield, VA, USA.

Tohoku University Inui Lab. 2022. BERT base Japanese (IPA dictionary). `https://huggingface.co/cl-tohoku/bert-base-japanese-char` (Accessed: 2022-12-18).

Joe. H. Ward Jr. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(31):236–244.

Haruhiko Kindaichi. 2010. *Japanese Language*. Tuttle Publishing, North Clarendon, VT, USA.

Nelson F. Liu, Gina-Anne Levow, and Noah A. Smith. 2018. Discovering phonesthemes with sparse regularization. In *Proceedings of the 2nd Workshop on Subword/Character Level Models*, pages 49–54, New Orleans, LA, USA.

Margaret Magnus. 2001. *What's in a Word? Studies in Phonosemantics*. Ph.D. thesis, Norwegian University of Science and Technology, Trondheim, Norway.

Susan Krupa McCune. 2011. *Exploring the combinatory effects of phonesthemes in brand naming*. Ph.D. thesis, California State University, Long Beach, CA, USA.

Katya Otis and Eyal Sagi. 2008. Phonaesthemes: A corpus-based analysis. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, pages 65–70, Austin, TX, USA.

Susan Janeen Parault and Paula J. Schwanenflugel. 2006. Sound-symbolism: A piece in the puzzle of word learning. *Journal of Psycholinguistic Research*, 35(4):329–351.

Plato. 1998. *Cratylus*. Hackett Publishing, Cambridge, MA, USA. translated by C. D. C. Reeve.

Plato. 1999. *Cratylus*. Project Gutenberg. `https://www.gutenberg.org/ebooks/1616`, translated by Benjamin Jowett (Accessed: 2022-12-18).

Åsa Abelin. 1999. *Studies in Sound Symbolism*. Ph.D. thesis, University of Gothenburg, Gothenburg, Sweden.

W. von Humboldt. 1836. *'On Language': On the Diversity of Human Language Construction and its Influence on the Mental Development of the Human Species*. Cambridge University Press, Cambridge, England, UK.

A Appendix

This section presents all of the phonesthemic clusters discovered in Section 5, where we applied the proposed method to 2,136 Kanji characters.

In Tables 7 and 8, the 100 clusters represented by the first consonant and the 33 clusters represented by the second consonant are listed in order of their representative consonant and z-score. The z-scores in the tables represent the test statistic for the population proportion test. In each phonesthemic cluster, Kanji characters having the representative consonant are presented in bold.

Table 7: Phonesthemic clusters represented by the first consonants and their semantic labels.

Phonesthemic Cluster	Representative Consonant	Proportion of Kanji Containing Representative Consonant	z-score	Semantic Label
価 果 均 献 値 **貢 勲 功** 績	k	7/ 9	4.311	feat
顧 苛 責 焦 煩 **甘 窮 困 苦** 辛 貧	k	6/11	2.845	—
講 教 校 究 学 **研** 祉 術 療	k	5/ 9	2.649	content
掘 坑 鉱 墾 拓 **耕** 牧 稲 隆	k	5/ 9	2.649	—
響 傾 向 影 調 **覚 感** 抑 精	k	5/ 9	2.649	sound property
閣 官 監 将 臣 **揮** 督 令	k	4/ 8	2.106	management
厚 深 浅 薄 余 **強 軽 堅** 弱 重 **固**	k	5/11	2.093	firm
厳 激 緩 快 急 **極** 著 微	g	3/ 8	4.092	acute
偽 痴 狂 盲 仮 **誤 疑** 迷	g	3/ 8	4.092	doubt
言 語 音 声 字 **号** 呼 称 名 番	g	3/10	3.511	language unit
郡 国 県 市 州 **郷 街** 都 府 里 村 町	g	3/12	3.069	administrative district
芸 劇 舞 美 演 **画** 作 像 絵 写 映 漫 撮 俳	g	3/14	2.716	representation
該 露 納 柄 相 **我** 荷 己	g	2/ 8	2.506	consciousness
獄 拷 囚 縛 拘 逮 拉 捕 虜	g	2/ 9	2.284	restraint
義 党 派 轄 理 **議** 委 協 盟	g	2/ 9	2.284	organization
技 拳 柔 剣 刀 **具** 棒 械 器	g	2/ 9	2.284	instrumentality
犠 牲 冥 禍 厄 **餓** 飢 愁 貪	g	2/ 9	2.284	sacrifice
岸 頂 畔 麓 峡 **岳** 渓 峠 岬 峰	g	2/10	2.093	natural elevation
供 恵 栄 寿 介 **丸 御** 援 助 衛	g	2/10	2.093	activity
宰 催 試 施 践 **載 紹** 提 搭 掲 披 導 搬	s	7/13	2.807	act
次 始 初 継 続 **終 先** 完 末	s	5/ 9	2.460	happening
清 静 聖 浄 鎮 **神 仙** 宮 天	s	5/ 9	2.460	change
象 処 措 候 案 **策 仕** 応 命	s	5/ 9	2.460	measure
裁 審 判 批 評 **算 察 測** 計 検 **査** 験	s	6/12	2.374	wisdom
鮮 濃 暗 汚 汽 **摂 触 染** 淡 喫 **掃 織** 吸	s	6/13	2.135	artifact
准 準 従 添 副 **随** 伴 沿 倣 逐	z	4/10	4.174	attendant
賊 稼 遊 敵 盗 **陣 侍** 忍	z	3/ 8	3.452	activity
純 指 管 統 総 **充** 整 備	z	2/ 8	2.047	control

Phonesthemic Cluster	Representative Consonant	Proportion of Kanji Containing Representative Consonant	z-score	Semantic Label
宜 遷 括 旦 般 **順** **徐** 頻	z	2/ 8	2.047	—
短 **長** **中** 少 小 **大** **低** 半 高	t	5/ 9	4.333	size
堆 **蓄** **貯** **騰** 旺 乏 湿 燥 漠 沃	t	4/10	2.981	condition
停 **滞** 休 止 留 **駐** **屯** 遣 泊 憩 睦	t	4/11	2.738	act
畜 匹 犬 猫 獣 **虫** **鳥** 魚	t	3/ 8	2.438	animal
糖 **窒** 蜜 酸 硫 **炭** 塩 油	t	3/ 8	2.438	compound
張 **貼** 割 擦 裂 **塗** 削 剥	t	3/ 8	2.438	stick
追 引 巻 押 越 **超** **突** 衝	t	3/ 8	2.438	force
秩 慈 淑 朴 妙 **泰** **貞** 潔 寧	t	3/ 9	2.183	morality
等 **他** 諸 各 同 **当** 翌 以 全 本	t	3/10	1.961	—
呈 **迭** 遂 嘱 繕 **捗** 賦 庸 寡 款	t	3/10	1.961	—
台 **団** **土** 落 所 **道** 場 地 路 域 区 駅 線 境 界 点	d	4/16	5.789	topographic point
堂 **洞** 院 寺 塔 坊 亭 楼	d	2/ 8	4.093	building
壇 **殿** 廷 典 威 誉 儀 宗 礼	d	2/ 9	3.806	activity
第 唯 最 単 複 **独** 自 彼 専 共 主 特	d	2/12	3.158	entirely
唾 血 尿 恥 汗 **瞳** 涙 耳 眉 虹 肌 髪	d	2/12	3.158	body covering
打 送 放 投 転 **動** 躍 活 駆 機 車 走 馬	d	2/13	2.990	turn
妊 胎 娠 姻 婚 **乳** 篤 酪 縫 癒	n	2/10	4.427	marriage
夏 冬 秋 春 週 **日** **年** 月 季 旬	n	2/10	4.427	time period
事 業 務 役 職 **任** **農** 商 工 働 労	n	2/11	4.180	duty
剛 豪 融 溶 乾 **軟** 洪 硬 滑 勾 **粘** 塑	n	2/12	3.964	—
女 **男** 子 婦 夫 妻 娘 兄 姉 弟 妹 親 父 母	n	2/14	3.599	woman
埼 茨 栃 畿 潟 阜 垣 曽 伎 岐 **那** 璃 瑠 縄 弥 **奈** 菜	n	2/17	3.170	—
悩 驚 泣 嘆 悦 笑 喜 悲	n	1/ 8	2.334	feeling
可 諾 益 済 届 **認** 申 許	n	1/ 8	2.334	permission
嬢 媛 姫 刹 婆 **尼** 僧 侶	n	1/ 8	2.334	woman
二 一 三 四 九 八 七 五 六	n	1/ 9	2.156	digit

Phonesthemic Cluster	Representative Consonant	Proportion of Kanji Containing Representative Consonant	z-score	Semantic Label
南 北 西 東 欧 韓 仏 英 米	n	1/ 9	2.156	cardinal compass point
能 素 実 格 質 材 品 物 気 風	n	1/10	2.003	artifact
肉 骨 脂 皮 殻 牙 歯 毛 矯 爪	n	1/10	2.003	connective tissue
脳 肝 腎 臓 腸 胃 肺 枢 核 炉	n	1/10	2.003	internal organ
弐 壱 渦 蚊 蛍 昨 昔 某 謎 闇	n	1/10	2.003	digit
奉 執 司 憲 宣 **布** **敷** 垂	h	3/ 8	2.864	artifact
符 **幣** 客 貨 券 **票** 駒 札 株	h	3/ 9	2.597	currency
俸 遇 酬 雇 璽 **罷** 劾 錮 賂 肖 **訃**	h	3/11	2.161	action
斑 骸 痕 跡 紋 **墳** **碑** 墓 陵 崖 丘	h	3/11	2.161	natural elevation
把 **蜂** **奮** 握 掌 勃 搾 嚇 穫 刈 **払** 絞 駄 頓 喝 蹴 捻	h	4/17	2.142	activity
避 逸 脱 逃 兼 **併** **排** **廃** 辞 含 除 散 消 滅 去 亡 退 了	h	4/18	2.009	act
怖 忘 恐 畏 憧 **飽** 萎 痩 臆 眺 **疲** 諦 衰 耗 慌 **弊** 綻 挫	h	4/18	2.009	—
扉 **壁** 窓 柱 柵 **塀** 鍵 錠 廊 架 鎖 塞	h	3/12	1.980	device
沸 湧 鳴 吐 吹 **噴** **拍** 繰 叫 跳 瞬 秒	h	3/12	1.980	utter
雌 雄 巣 餌 卵 **哺** 昆 脊 臼 穂 **帆** 貝 藻 酵 菌 **胞** **泡** 滴 晶 豆	h	5/25	1.975	foodstuff
売 配 頒 販 購 **買** **貿** 需 輸 逓 郵	b	3/11	3.570	commerce
盆 **瓶** 皿 鉢 缶 袋 箱 枠	b	2/ 8	2.739	container
部 面 方 側 回 **分** 度 会 合 間	b	2/10	2.310	distance
罵 殴 叱 怒 憤 **侮** 慄 慨 嘲 蔑	b	2/10	2.310	discourtesy
尾 背 腹 胸 腰 **鼻** 首 頭 尻 口	b	2/10	2.310	body part
冒 危 緊 脅 侵 **防** 捜 探 偵 警	b	2/10	2.310	policeman
暴 **妄** 虐 淫 鬱 酷 惨 険 疾 痛	b	2/10	2.310	miserable
伐 征 討 填 却 撤 渉 訟 訴 締 **謀**	b	2/11	2.136	group action
盤 序 壌 礎 姓 **苗** 拠 領 籍 契 約	b	2/11	2.136	book
武 **文** 世 代 治 **民** 住 政	m	3/ 8	5.723	geological time

Phonesthemic Cluster	Representative Consonant	Proportion of Kanji Containing Representative Consonant	z-score	Semantic Label
夢 幻 魂 霊 怪 **魔** 妖 呪	m	2/ 8	3.649	spirit
紡 蚕 絹 繭 糸 **麻** **綿** 繊 桑 藍	m	2/10	3.153	shrub
必 須 要 依 寄 **問** 負 課 臨 求 **望** 待 補	m	2/13	2.619	—
照 景 灯 図 譜 **目** 鑑 鏡 綱 眼 **幕** 旗 傘	m	2/13	2.619	artifact
憾 惧 唆 摘 戒 匿 蔽 粛 謹 慎 **昧** **慢** 遜	m	2/13	2.619	activity
用 効 利 使 注 **有** 無 収 獲 採 **与** 取 加 承 受 得	y	3/16	3.619	accept
誘 奨 勧 薦 推 **諭** 促 励	y	2/ 8	3.580	rede
踊 **謡** 唄 歌 唱 奏 詞 詩 楽 曲	y	2/10	3.089	chant
欲 悔 寂 惜 乞 請 願 祈 誓 懸 **預** 賭 借 貸 譲 託	y	2/16	2.174	—
羅 **麗** 唐 漢 呉 **籠** 揚 束	r	3/ 8	4.069	artifact
瞭 凡 悠 確 剰 **累** **零** 僅 裕 遡 暫 漸	r	3/12	3.049	—
烈 勇 敏 恭 孝 **廉** 賢 謙 簡 容 **頼** 尋 誠 忠	r	3/14	2.697	—
角 翼 針 標 玉 **鈴** **輪** 環	r	2/ 8	2.490	machine
恋 仲 友 愛 模 **旅** 宿 師	r	2/ 8	2.490	sexual desire
連 **絡** 携 関 係 結 接 雑 密	r	2/ 9	2.268	change
老 稚 若 幼 児 **齢** 童 才 歳	r	2/ 9	2.268	time of life
帳 簿 箋 抄 謄 **欄** **漏** 喪 抹 抽	r	2/10	2.077	written record
賄 拾 伺 拭 漂 浮 洗 覆	w	1/ 8	6.529	freewheel
空 海 陸 邦 洋 **和** 星 宙 泳 球 塁	w	1/11	5.520	region
解 釈 説 論 談 **話** 証 識 知 告 報 示 表 録 記 述	w	1/16	4.510	—

Table 8: Phonesthemic clusters represented by the second consonants and their semantic labels.

Phonesthemic Cluster	Representative Consonant	Proportion of Kanji Containing Representative Consonant	z-score	Semantic Label
黒 白 赤 青 紫 **緑** 紅 黄	k	4/ 8	3.568	chromatic color
服 衣 靴 帽 飲 **食 浴** 酒 醸 粧 **濯**	k	4/11	2.727	clothing
角 翼 玉 針 標 鈴 環 輪	k	3/ 8	2.429	machine
搾 嚇 蜂 奮 勃 **穫** 刈 絞	k	3/ 8	2.429	extort
憶 択 答 考 思 **郭** 想 構	k	3/ 8	2.429	belief
易 略 難 便 通 **激** 緩 快 急 厳 **極** 著 微	k	4/13	2.316	condition
植 殖 育 養 培 栽 飼 丹 肥 豊 **拓** 掘 坑 鉱 墾 **牧** 稲 耕 隆	k	5/19	2.174	cultivate
複 独 単 自 彼 **特** 専 共 主	k	3/ 9	2.174	entirely
滴 穀 晶 泡 豆 **麦** 粉 菓 粒	k	3/ 9	2.174	grain
寂 惜 欲 悔 乞 請 願 祈 誓	k	3/ 9	2.174	invite
忘 恐 怖 畏 憧 **臆** 飽 萎 痩 眺 諦 疲 衰 耗 慌 綻 挫 弊 朽 腐	k	11/90	0.428	—
効 利 使 用 注 **獲** 無 有 収 採 取 加 与 承 受 **得**	k	2/16	0.216	accept
惑 拐 踪 陥 堕 **縛** 墜 奴 隷 囚 **獄** 拘 拷 逮 拉 **辱** 捕 虜 恩 褒	k	4/34	0.177	—
桃 梨 柿 梅 桜 **菊 竹** 滝 柳 松 杉 猿 鬼 蛇 竜 亀 虎 鶴 熊 鹿	k	6/56	-0.025	—
撲 棋 碁 俵 升 斗 埼 茨 栃 畿 潟 阜 垣 曽 伎 岐 璃 瑠 縄 弥	k	1/23	-0.999	—
携 連 関 係 絡 **結 接 雑 密**	t	4/ 9	5.083	change
割 擦 裂 張 貼 塗 削 剥	t	3/ 8	3.936	stick
一 二 三 四 九 **八 七** 五 六	t	3/ 9	3.630	digit
刷 筆 稿 刊 版 **閲** 書 描 読	t	3/ 9	3.630	text
抜 切 掛 込 組 **決** 選 編 予 定	t	3/10	3.367	statement
実 質 素 能 格 **物** 材 品 気 風	t	3/10	3.367	artifact
立 置 持 存 産 **出** 生 入 成 行 **発**	t	3/11	3.137	act
窟 穴 隙 孔 栓 **圧** 槽 膜 筒 液 蓋 胴	t	3/12	2.933	opening
蜜 窒 糖 酸 硫 塩 炭 油	t	2/ 8	2.395	compound

Phonesthemic Cluster	Representative Consonant	Proportion of Kanji Containing Representative Consonant	z-score	Semantic Label
越 引 巻 押 追 **突** 超 衝	t	2/ 8	2.395	force
秩 慈 淑 朴 妙 **潔** 泰 貞 寧	t	2/ 9	2.177	morality
逸 **脱** 逃 避 兼 併 辞 排 廃 含 **滅** 除 散 消 去 亡 退 了	t	3/18	2.051	act
罵 殴 叱 怒 憤 **慄** **蔑** 慨 嘲 侮	t	2/10	1.989	discourtesy
鬱 虐 暴 淫 妄 **疾** 酷 惨 険 痛	t	2/10	1.989	miserable
夏 冬 秋 春 週 **日** **月** 年 季 旬	t	2/10	1.989	time period
蚕 **絹** **繭** 紡 糸 **繊** **藍** **綿** 桑 麻	N	6/10	3.098	shrub
遷 **亘** **般** 宜 括 **順** **頻** 徐	N	5/ 8	2.947	—
原 **森** 沢 谷 野 **林** **園** **山** 島	N	5/ 9	2.609	region
進 **信** **伝** 流 交 **運** **展** 情 集 達	N	5/10	2.315	group action
准 **準** **添** 副 従 **伴** **沿** 倣 随 逐	N	5/10	2.315	attendant
敏 勇 烈 恭 孝 **賢** **謙** **廉** **簡** 容 **尋** 頼 誠 忠	N	6/14	2.076	—
艦 **船** **桟** 舟 艇 **舷** 隻 舶	N	4/ 8	2.070	vessel
煩 顧 責 焦 苛 **甘** **辛** **困** 窮 苦 **貧**	N	5/11	2.054	—