Predicting the structure of a lexical environment from properties of verbal working memory

Yuki Tanida [a, b] & Satoru Saito [c]

[a] *Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University, Osaka, Japan*
Tel: +81 72 252 1161 (extension 4419)
E-mail: yuki.tanida@hs.osakafu-u.ac.jp
Address: 1-1 Gakuencho, Naka-ku, Sakai City, Osaka, 599-8531, Japan.

[b] *Japan Society for the Promotion of Sciences, Tokyo, Japan*
Address: 5-3-1 Kojimachi Business Center, Kojimachi, Chiyoda-ku, Tokyo, 102-0083, Japan

[c] *Graduate School of Education, Kyoto University, Kyoto, Japan*
Tel: +81 75 753 3067
E-mail: saito.satoru.2z@kyoto-u.ac.jp
Address: 36-1, Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan

Correspondence to:
Yuki Tanida
Tel: +81 72 252 1161 (extension 4419)
E-mail: yuki.tanida@hs.osakafu-u.ac.jp
Address: Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University, 1-1 Gakuencho, Naka-ku, Sakai City, Osaka, 599-8531, Japan.

Keywords: corpus analysis, imageability, phonotactic frequency, accent regularity, working memory, long-term memory, arbitrariness of the sign

Tanida, Y., & Saito, S. (2022). Predicting the structure of a lexical environment from properties of verbal working memory. *Cognitive Science*, *46*, e13181. https://doi.org/10.1111/cogs.13181

Abstract

We analysed a Japanese lexical database to investigate the structure of the lexical environment based on the hypothesis that the lexical environment is optimized for the functioning of verbal working memory. Our prediction was that, as a consequence of the cultural transmission of language, low-imageable meanings tend to be represented by frequent phonological patterns in the current vocabulary rather than infrequent phonological patterns. This prediction was based on two findings of previous laboratory studies on verbal working memory. (a) The quality of phonological (phonemic and accent) representations in verbal working memory depends on phonological regularity knowledge; therefore, short-term phonological representations are less robust for words with infrequent phonological patterns. (b) Phonological representations are underpinned by contributions from semantic knowledge; therefore, phonological representations of highly imageable words are more robust than those for low-imageable words. Our database analyses show that nouns with less imageable meanings tend to be associated with more frequent phonological patterns in Japanese vocabulary. This lexical structure can maintain the quality of phonological representations in verbal working memory through contributions of semantic and phonological regularity knowledge. Larger semantic contributions compensate for the less robust phonological representations of infrequent phonological forms. The quality of phonological representations is preserved by phonological regularity knowledge when larger semantic contributions are not expected.

1. Introduction

Recent studies analysing linguistic corpora (for a review of this method, Christiansen & Monaghan, 2016) have shown that language is structured for usability for humans (e.g., Christiansen & Chater, 2008; Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Gibson, Futrell, Piantadosi, Dautriche, Mahowald, Bergen, & Levy, 2019; see also Zipf, 1949), particularly in terms of the learnability (Dautriche, Mahowald, Gibson, & Piantadosi, 2017; Kelly, 1992; Lewis & Frank, 2016; Monaghan, Christiansen, & Fitneva, 2011; Monaghan, Shillcock, Christiansen, & Kirby, 2014, see also Schmidtke, Conrad, & Jacobs, 2014) and communication efficiency (Lewis & Frank, 2016; Piantadosi, Tily, & Gibson, 2011) of words. The present study predicted the usable lexical structure from the mechanisms of verbal working memory, which has not been widely explored in this field but is a fundamental cognitive function for the learning and communication of words, with a focus on the Japanese lexical structure. We start this paper with a brief introduction to verbal working memory and Japanese phonology.

Verbal working memory (for a review, see Oberauer et al., 2018) is a memory function that underpins the transient retention of a sequence of linguistic elements in language-based cognitive activities, such as the recognition and production of words (e.g., Allen & Hulme, 2006; Majerus, 2013) and word learning (Baddeley, Gathercole, & Papagno, 1998; Gathercole, 2006; Gupta, 2009; Gupta & Tisdale, 2009a; 2009b; Majerus & Boukebza, 2013; Majerus, Perez, & Oberauer, 2012; Majerus, Poncelet, Elsen, & van der Linden, 2006; Majerus, Poncelet, Greffe, & van der Linden, 2006; but see also Melby-Lervåg, Lervåg, Lyster, Klem, Hagtvet, & Hulme, 2012). Various conceptual and computational models have been proposed for this memory function (e.g., Burgess & Hitch, 1992; 1999; 2006 for computational models of the phonological loop). Specifically, most of the computational models of verbal working memory have

focused on mechanisms of serial order processing (e.g., Botvinick & Plaut, 2006; Farrell, 2012; Gupta, 2009; Gupta & Tisdale, 2009a; 2009b; Henson, 1998; Lewandowsky & Murdock, 1989; Page & Norris, 1998; 2009; for a review, Hurlstone, Hitch, & Baddeley, 2014).

An important property of verbal working memory is that the quality of phonological[1] representations in working memory depends on long-term knowledge of the statistical distribution of phonological information in the linguistic environment (e.g., Gupta & Tisdale, 2009a; 2009b; Hartley & Houghton, 1996; for a collection of articles from this field, see Thorn & Page, 2008; for a review, see Saito, Nakayama, & Tanida, 2020). The effect of phonotactic frequency is an example of a phenomenon that reflects long-term serial order knowledge in the phonemic domain. This effect is observed predominantly in the oral recall of auditorily presented verbal items; recall performance is higher for nonwords composed of phonotactic combinations that appear frequently in a particular language than for nonwords composed of infrequent phonotactic combinations. Specific cases of the effect are known as the bi-phone frequency effect in English and *the bi-mora*[2] *frequency effect* on immediate serial recall in Japanese (for studies of English, see Gathercole, Frankish, Pickering, & Peaker, 1999; Thorn, Gathercole, & Frankish, 2005, for studies of Japanese, see Tanida, Nakayama, & Saito, 2019; Tanida, Ueno, Lambon Ralph, & Saito, 2015a) and single-

[1] We use terminology typical in the research field of phonological/verbal working memory. Phonological information on word forms has *phonemic* and *accent* aspects. The phonemic aspect is composed of consonants or vowels. The accent aspect is composed of within-word pitch patterns in the case of Japanese. Therefore, the term *phonological* is an abstract concept more than terms *phonemic* and *accent*. The term *phonological* is used when both phonemic and accent aspects are described or when we do not specify either phonemic or accent aspects.

[2] A mora is a subsyllabic and basic phonological unit in Japanese (Kubozono, 1995; McQueen, Otake, & Cutler, 2001). A mora is composed of a vocalic nucleus (V), a vowel with an onset (together, CV or CCV), a nasal consonant (N) in syllabic coda position, and a geminate consonant (Q) or long vowel (R).

item repetition (for a study of English, see Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997, for a study of Japanese, see Tanida, Ueno, Lambon Ralph, & Saito, 2015b).

Japanese involves another type of statistical distribution in the prosodic domain, i.e., accent regularity (Sakono, Ito, Fukuda, & Fukuda, 2011; Tanida et al., 2015a; 2015b; Ueno Saito, Saito, Tanida, Patterson, & Lambon Ralph, 2014). Japanese is a pitch-accent language whereby the pitch pattern can discriminate words. For example, the bi-mora sequence /a-me/ means 'candy' when stated with a low-high pitch pattern but 'rain' when said with a high-low pitch pattern. In the standard theory of Japanese accents (Kindaichi, 2001), accent type is categorized by the position of the pitch drop that emerges under the following three rules: (1) pitch shifts across the first and second morae (if the pitch of the first mora is high, the pitch of the second mora is low and vice versa); (2) once pitch drop occurs, pitch never rises; and (3) pitch does not necessarily drop within a word (i.e., flat). Because N-mora words have N positions for potential pitch drop, there are N + 1 accent types (see Table 1). Importantly, the occurrence frequency varies among accent types as presented in Table 1. Tanida and colleagues found an effect of Japanese accent frequency on nonword immediate serial recall (Tanida et al., 2015a) and on single nonword repetition (Tanida et al., 2015b). They specifically reported that more accent errors occurred for nonwords presented with infrequent accent types than for nonwords presented with frequent accent types. Thus, the verbal working memory performance of Japanese speakers is sensitive to the statistical distribution of both phonemic and accentual aspects of phonology. It is assumed that the regularity of phonological aspects secures the quality of phonological representations in verbal working memory.

[Table 1]

The quality of the phonological representation of verbal sequences in verbal working memory also depends on the richness of feedback from the activated semantic representation associated with the verbal sequences. It is well known that activation of semantic representation improves the retention of phonemic sequences in verbal working memory (e.g., Gupta & Tisdale, 2009b; Schweickert, 1993), as well as the recognition and production of verbal items (e.g., Luce, Goldinger, Auer, & Vitevitch, 2000; Vitevitch & Luce, 1999; for a model combining activities of retention, recognition, and production, see Ueno, Saito, Rogers, & Lambon Ralph, 2011; Ueno et al., 2014). This improvement has been indicated by better performance on phonological short-term memory tasks with real words than with nonwords (e.g., Jeffries, Frankish, & Noble, 2009; Hulme, Maughan, & Brown, 1991; Thorn et al., 2005; also, on recognition/auditory initial phonemic detection, Rubin, Turvey, & van Gelser, 1976; on production/letter-string naming, e.g., Forster & Chambers, 1973; on repetition, Roy & Chiat, 2004; Vitevitch & Luce, 1999). In particular, the semantic contribution to the retention of phonemic sequences is greater for high-imageable words than for low-imageable words, a phenomenon known as *the imageability effect* (Tyler, Voice, & Moss, 2000; Ueno, 2012; Ueno et al., 2014).

Although evidence is still limited, the influence of semantic contributions (e.g., imageability) on the retention of accent patterns in Japanese has been reported. Ueno (2012; Ueno et al., 2014) showed that in immediate serial recall and single-word repetition tasks, Japanese adults recalled the words with an accent type that was the same as the presented type (correct retention of the accent pattern) more for high-imageable word items than for low-imageable word items. Furthermore, Ueno et al.

(2014) reported that the imageability effect on the accuracy of accent pattern retention in immediate serial recall was largest for the most infrequent accent type.

In summary, the retention of phonological (phoneme and accent) representations in verbal working memory depends on the phonological regularity of the verbal sequences, which reflects phonotactic (bi-mora in Japanese) and accent frequency in the linguistic environment, and on the imageability of the meanings, which is associated with the richness of the semantic contribution.

Deducing from these properties of verbal working memory, a vulnerable case involves words using infrequent phonological patterns to represent less-imageable meanings. Such form-meaning mappings have fragile phonological representations in verbal working memory because neither phonological nor semantic knowledge can contribute sufficiently. Consequently, the probability of error for such words will be high in linguistic activities such as recognition, production, and learning. However, do we often face such problems? It has been suggested that language has been structured through cultural transmission (e.g., Kirby, Cornish, & Smith, 2008; Monaghan et al., 2014; Scott-Phillips & Kirby, 2010; Smith, Kalish, Griffiths, & Lewandowsky, 2008). For successful cultural transmission, words must be learnable and must satisfy communication efficiency requirements in noisy environments to survive. Therefore, words that are difficult to learn, recognize and produce correctly cannot linger on, and consequently, the current vocabulary will be less likely to retain words composed of such form-meaning mappings.

The present study investigated this possibility that the structure of the linguistic environment is sensitive to the quality of phonological representation in verbal working memory. We predict that in the current lexical environment, there are few less-imageable words that consist of infrequent phonological patterns. Nevertheless,

language needs to be able to hold low-imageable meaning with probabilistically distributed phonological patterns. A possible manner in which words with infrequent forms or low-imageable meanings can survive might be as follows. Phonological forms whose short-term representation is fragile (e.g., forms with infrequent phonotactic and accent patterns) are assigned to meanings with rich semantic contributions (e.g., high-imageable meanings). Conversely, meanings providing poor semantic contributions (e.g., low-imageable meanings) are accompanied by phonological forms with robust short-term phonological representation (e.g., forms with frequent phonotactic and accent patterns). With such mapping characteristics in the vocabulary space, the quality of phonological representation in verbal working memory can be kept at a certain level by contributions of either semantic or phonological knowledge or both. Therefore, it can be assumed that the lexical environment has been structured across generations in such a way that there is a negative correlation between imageability and phonological regularity.

In analyses of the hypothesized form-meaning mappings in vocabulary, we considered multiple psycholinguistic variables potentially affecting verbal working memory and/or correlating with imageability and phonotactic regularity. We first investigated relationships of imageability with phonotactic or accent regularity by Spearman's rank correlation analysis and Brunner-Munzel test, respectively. Finding the predicted negative correlations between imageability and phonological regularity would indicate that the Japanese vocabulary structure is able to maintain the quality of phonological representation in verbal working memory through the contributions of phonological and semantic knowledge in a mutual compensatory manner. Then, we examined whether the relationships between imageability and phonological regularity, if present, were kept in regression analyses even with controlling covariates. This

analysis tested whereby the observed relationships could not be explained by a spurious correlation emerging across available variables.

2. Dataset preparation

Our target variables were imageability, average bi-mora token frequency, and accent regularity. We also considered word length (the number of phonemes), average single-mora token frequency, word frequency, neighbourhood size, and etymological word type as control variables. We employ etymological word type because Railly & Kean (2007) reported its correlation to other variables including imageability in English. To prepare the dataset, we employed the largest Japanese word frequency database, *the short unit word list of BCCWJ (Balanced Corpus of Contemporary Written Japanese),* established by the National Institute for Japanese Language and Linguistics (version 1.1, NINJAL, 2015). All words included in this database are composed of the shortest morphological unit defined by NINJAL, named the *short unit*.[3] Figure 1 shows a schematic image of our dataset preparation process.

[Figure 1]

The accent types (see Table 1), phonemic pronunciations, and etymological word types for all entries in the short unit word list were retrieved from UniDic (version

[3] There is a difficulty in defining a monomorphemic word unit for Japanese. For example, the word '気持ち' /kimochi/ (meaning feeling/sensation/intention/mood in English) can be further decomposed into shorter units of '気' /ki/. However, if '気持ち' is recognized as a compound word or multimorphemic word, and is decomposed into two units of '気' and '持ち', token frequency of '持ち' includes occurrence of '持ち' in a context of '私はその本を*持ち*…' (I *have* the book…) and occurrence of '持ち' in '気持ち'. This is not natural because '持ち' in '気持ち' does not work as a verb "have". Therefore, NINJAL defines (and the short unit word list of BCCWJ registers) '気持ち' as a word composed of the shortest morphologically meaningful unit.

2.3.0, NINJAL, 2018). Because some words allow multiple accent types[4], we retrieved, for such words, an accent type registered as the most typical type in UniDic. Words were categorised into four etymological types. Native Japanese words were inherited from the old Japanese language. Sino-Japanese words originated from Chinese. These two types are considered traditional Japanese words. Loanwords are derived mainly from Western languages and are a relatively new word type. Most of them appeared after the 19th century. Hybrid words are composed of parts of multiple etymological types. An example is the word "円安" /e-n-ya-su/, meaning "weaker yen". The component "円" /e-n/ represents Japanese yen, which is a Sino-Japanese word, and "安" /ya-su/ derives from the native Japanese word "安い" /ya-su-i/, meaning "cheap".

The imageability values were retrieved from another largest psycholinguistic database, the *NTT database series: Lexical properties of Japanese* (Sakuma et al., 2005).[5]

Phonotactic regularity was represented as an average of the log-transformed token frequency of the bi-mora components in each word. First, we calculated the token frequency of every possible bi-mora from word frequency in the short unit word list of BCCWJ.[6] The bi-mora frequency for each word was defined as the average of the log-transformed (base-10) bi-mora token frequencies of all bi-morae in each word.[7] We

---

[4] For example, a word '映画' /eiga/ (meaning 'movie' in English) can be pronounced with a flat type accent or type 1 accent. The dictionary registers flat type as the first for the word.

[5] The database covers two types of imageability: verbal imageability, evaluated based on auditory presentation, and orthographic imageability, evaluated based on orthographic presentation. Both types of imageability are evaluated by a seven-point Likert scale. We employed verbal imageability. Note that some word forms are associated with multiple meanings; thus, reported rating values are based on the meaning that participants reported in the imageability ratings. We retrieved imageability values based on the rating for the target meaning.

[6] For example, the token frequency of the bi-mora /ishi/ is the sum of the frequencies of words containing /ishi/, e.g., '医師' /ishi/ (doctor) and '名刺' /meishi/ (name card).

[7] For example, for the word /shinrigaku/ (meaning 'psychology'), the composing bi-morae are /shin/, /nri/, /riga/, and /gaku/.

added one to the token frequency of each bi-mora because the frequency for some bi-mora tokens was zero, for which we could not calculate the logarithmic value. In addition, we calculated the token frequency of each single mora and the average token frequency of each mora as done for the token frequency of bi-mora components.

For the neighbourhood size of each word, we counted the number of words that share a syllabic structure in the short unit word list of BCCWJ. Neighbours were defined as words that shared the same number of syllables; the same syllable position of the accent (pitch-drop); and all phonemes if a phoneme was added, omitted, or substituted.[8] We identified the syllable position of the accent in relation to the end. Neighbourhood size was log-transformed with base-10 to approximate a normal distribution as much as possible. We added 0.5 before the log-transformation because the neighbourhood size for some words was zero. The long vowel /R/ was considered to be the same vowel as the preceding vowel when calculating bi-mora frequency, single-mora frequency, and neighbourhood size.

Accent regularity was assigned following a Japanese accent assignment rule based on phonological structure regardless of word length (Kubozono, 2006). According to this rule, there are two main types of Japanese accents: flat (nonaccented) and accented. The former represents the flat type, and the latter corresponds to the type N accent patterns in Table 1 (N represents the mora position with pitch drop counted from the onset). The flat accent is most frequent in Japanese vocabulary. It is still not clear how to distinguish flat nouns from accented nouns, whereas there is a rule for assigning accent position among accent types (Kubozono, 2006): accent (pitch drop) is assigned to a *syllable* containing the antepenultimate *mora*. In accordance with this rule,

[8] We translated the phonemic word forms registered in the short unit word list of BCCWJ into syllable-based structures by defining the heaviest syllabic structure as CVVC, according to Kubozono (1995).

we categorized all noun entries with more than two morae into three types: *flat pattern*, *regularly accented pattern*, and *irregularly accented pattern*.[9] It is impossible to apply the accent rule to one- and two-mora words. For two-mora words, accent regularity was coded as regularly accented type for type 1 accent words and as irregularly accented type for type 2 accent words because the type 1 accent is more frequent than the type 2 accent (see Table 1). One-mora words were removed from the analysis dataset because a one-mora word allows only two levels of accent type (flat and type 1; see Table 1), and it is impossible to assign bi-mora frequency to one-mora words.[10]

Finally, our dataset contained 18,262 common noun entries, each composed of multiple morae with all variables to be analysed. Table 2 provides abbreviated labels for the variables used in this paper. Note that because the imageability ratings in the NTT database were conducted only for entries with sufficient familiarity (four or more than four from one to seven ratings) and we selected words with imageability values for our analyses, our prepared database can be regarded as a comprehensive collection of Japanese words that speakers use regularly. We uploaded the Perl code for dataset preparation and the R code for statistical analyses to the Open Science Framework repository (https://osf.io/z7hj3/).[11]

[Table 2]

3. Analysis of the whole dataset

[9] For example, the four-mora and two-syllable word '別荘' /be-Q-ˈso-o/, meaning 'villa', is categorized as an irregularly accented word, and the four-mora and three-syllable word '骸骨' /ˈga-i-ko-tsu/, meaning 'skeleton', is categorized as a regularly accented word. Here, each mora was separated by a hyphen. An apostrophe appears at the front of an accented mora.

[10] The removal of these words had a negligible effect on the analysis because one-mora nouns represent only 0.43% of all noun entries in the short unit word list of BCCWJ (437 of 101,603 common nouns), of which accent type was available for 360 common nouns.

[11] The original database cannot be uploaded due to copyright restrictions.

Figure 2 shows the relationships between all continuous variables. Figure 3 shows the relationship between imageability and accent regularity. We here only focused on the statistical results relating to the relationship between imageability and phonological regularity.

[Figure 2]

[Figure 3]

We performed Spearman's rank correlation analyses for six continuous variables. Because we had a total of 15 correlations, the significance level was set as 0.0033 (= 0.05/15) by Bonferroni correction. We found a predicted significant negative effect (r = -0.177) in the correlation of imageability with the bi-mora frequency. This correlation is small ($0.10 < r < 0.30$) according to Cohen (1992).

We performed three Brunner-Munzel tests to compare imageability between three types of accent regularity. Table 3 shows the results of the Brunner-Munzel tests and effect size (Cliff's delta). The significance level was set as 0.0167 (= 0.05/3) by Bonferroni correction. As we predicted, imageability was the highest for irregularly accented nouns and was higher for regularly accented nouns than for flat nouns. As shown in the effect sizes (Cliff's delta) between categories, the gap was particularly large between the most frequent accent type (flat) and the most infrequent irregularly accented pattern.

[Table 3]

Then, regression analyses evaluated the relationship between imageability and phonological regularity while controlling for other variables. We first performed an ordinary least square regression (OLSR) analysis. The dependent variable was standardized imageability. All other five continuous variables were standardized and projected for independent variables. Accent regularity and etymological word type were also projected to the model with dummy coding. The intercept was the level of irregularly accented patterns and Sino-Japanese words. The result of the OLSR model is shown in the left column of Table 4 with the variance inflation factor (VIF) for each variable. As the VIF for neighbourhood size shows a higher value (> 5), collinearity was problematic. Thus, we further performed two types of regression analyses to deal with such collinearity and render the result more confirmatory. One method applied the conventional approach of the same OLSR analysis but excluding the variable of neighbourhood size. The result is presented in the middle column of Table 4. We then performed a likelihood ratio test to compare a model with all variables (without neighbourhood size) and a model without the term of bi-mora frequency to investigate the need for the term of bi-mora frequency. The model with the term of bi-mora frequency showed a significantly better fit than the model without the term (see Table 4).

[Table 4]

The second approach used *supervised component generalized linear regression* (SCGLR, Cornu, Mortier, Trottier, & Bry, 2016) introduced by Tomaschek, Hendrix, and Baayen (2018). This method estimates betas involving all variables including neighbourhood size while reducing the impact of collinearity on the result. A detailed

explanation and the procedure of the SCGLR analysis, as well as the linearity check, are provided in *Supplementary Information A*. Briefly, this method applies a family of principal component regression and reduces the problem of collinearity by reducing information included in given data. As a result, the estimated beta for each variable will be biased toward zero and take a milder value. Such mild regression coefficients can be recognized as betas less reflecting the impact of collinearity. The estimated betas for each variable are shown in the right column of Table 4. There are two points to note, however. The standard error and *p value* are not available for the estimated betas of continuous variables. In addition, information reduction in data is not done for categorical variables, but standard errors and *p values* are available. Therefore, we recognized a significant coefficient for continuous variables when the coefficients were significant ($p < .05$) in the two OLSR analyses and the sign of the coefficient was kept across three models. For categorical variables, a significant coefficient was recognized when the coefficients were significant ($p < .05$) with the same sign across all three models. According to this criterion, bi-mora frequency showed a significantly negative coefficient. The term for the flat accent pattern had a significantly positive coefficient, but that for the regularly accented pattern did not.

4. Interim discussion

The Japanese vocabulary showed a predicted significant negative correlation between imageability and bi-mora frequency in both Spearman's correlation and regression analyses. The analysis of accent regularity showed imageability to be higher for nouns with irregularly accented nouns than for flat and regularly accented nouns in the Brunner-Munzel tests. However, this relationship disappeared when other variables were controlled in regression analyses. Rather, the regression models predicted lower

imageability for irregularly accented nouns relative to the other two accent types. One possible reason for this unpredicted result is that the accent type distribution in loanwords, whose imageability was the highest, was unique. Figure 3 presents the type frequency of each accent regularity category in the whole dataset as a function of etymological word type. Flat words are the most infrequent for loanwords, unlike other word types. This indicates the possibility that loanwords are not much integrated into traditional Japanese vocabulary. To explore these results further, we reanalyzed the dataset excluding loanwords through the same procedure of analyses as the whole database analysis. The dataset without loanwords had 15,365 common nouns.

[Figure 4]

5. Analysis of the dataset without loanwords

As shown in Figure 2, the Spearman's rank correlation between imageability and bi-mora frequency still showed a small level (Cohen, 1992) of the significant negative correlation (r = -0.185). The results of the three regression analyses are presented in Tables 4 as well as the likelihood ratio test to investigate the need for the term of bi-mora frequency in the OLSR model, similar to the analysis of the whole dataset. They again indicate significant-negative coefficients for bi-mora frequency.

Figure 3 and Table 3 indicate that the median imageability was significantly higher for irregularly accented nouns than flat nouns and regularly accented nouns, but the imageability of nouns with these two frequent accent types did not show a significant difference. The results of the three regression analyses presented in Tables 4 show that the term for regularly accented nouns had a significantly negative coefficient. The term for flat nouns did not reach a significant level, although the sign was negative.

6. Discussion

We found that nouns composed of less frequent phonotactic patterns tend to represent more imageable meanings in Japanese vocabulary. This negative correlation between imageability and phonotactic frequency was reliably detected across Spearman's correlation and regression analyses while controlling other variables in datasets with or without loanwords. This structure of the lexical environment is *able* to compensate for the quality in phonemic representations with the contribution of phonotactic or semantic knowledge. We propose that when phonemic representation is not robust due to smaller contributions of phonotactic knowledge (i.e., due to its phonotactically irregular form), semantic contributions tend to strongly support phonemic representations and vice versa.

For the relationship between imageability and accent regularity, we focus on the results of the dataset without loanwords since the accent structure of loanwords was regarded as the one that has not been sufficiently integrated in Japanese. The Brunner-Munzel tests revealed that irregularly accented nouns had significantly higher imageability than flat and regularly accented nouns. The negative correlation-like relationship between imageability and accent regularity in the lexical environment is assumed to reflect a mutually compensatory relationship of semantic and accent regularity knowledge, which can contribute to the preservation of the quality of accent representation in verbal working memory. However, a significant gap in imageability between irregularly accented nouns and flat nouns was not detected in regression analyses after controlling the related variables. The relationship of imageability with accent regularity might be relatively weaker than that with phonotactic regularity. A

potential factor fuzzifying the predicted negative correlation involving the serial position effect is discussed in *Supplementary Information B*.

The reciprocal, and thus systematic, relationship between the contribution of phonological and semantics knowledge seems to somewhat violate the arbitrariness of the sign (de Saussure, 1916/1973; Hockett, 1963). However, it might be reasonable to state that the reciprocal structure reported here is a result of the arbitrariness of the sign. The general principle allows any mappings between form and meaning and therefore creates a space for language users to select *useful* mappings. Meanwhile, although mapping freedom is constrained by various pressures on the cultural transmission/evolution of language, the pressure does not extinguish arbitrariness because of its benefit of making the lexical space sparser, leading to smooth semantic access and phonological form production (Gasser, 2004; Lambon Ralph, Moriarty, & Sage, 2002; Monaghan et al., 2014). Under this principle, the similarity of phonological forms does not correlate with the similarity of meanings, and words with similar phonological information are mapped, in semantic space, with some distance between them. Consequently, semantic competitors of the target word have less phonological overlap with the target. This cognitive benefit of the arbitrariness of the sign has enabled words with this property to survive cultural evolution and remain in the current lexical structure. In this light, we can understand the results of our OLSR models, which explain approximately 25% of the variance (see the adjusted R^2s in Table 4). It is assumed that our results reflect the balance between systematicity and arbitrariness in form-meaning mappings in Japanese.

Finally, it should be noted that we did not directly observe a causal relationship whereby verbal working memory constrains linguistic structure and similarly do not directly confirm that the structure of the Japanese linguistic environment truly

contributes to the quality of short-term phonological representation. Therefore, we must continue to scrutinize the existence of a direct causal relationship between verbal working memory and the linguistic environment. Additionally, the negative correlation between imageability and phonological regularity may be unique to Japanese. Japanese is an agglutinative language. A noun word is followed by a particle word. Since nouns are not inflected, the word form does not change across contexts. Therefore, it may be that relative to other types of languages such as fusional and isolating languages, the Japanese language more directly reflects working memory mechanisms in the word form, robustly realizing a negative correlation between word form and imageability.

Disclosure of interests. The authors report no conflicts of interest.
None of the data or materials for the analyses reported herein are available, and none of the analyses were preregistered.

References

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language, 55*(1), 64-88. doi: 10.1016/j.jml.2006.02.002

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158-173. doi: 10.1037//0033-295x.105.1.158

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: a recurrent neural network model. Psychological review, 113(2), 201−233. doi: 10.1037/0033-295X.113.2.201

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language, 31*(4), 429-460. doi: 10.1016/0749-596X(92)90022-P

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*(3), 55-5811. doi: 10.1037/0033-295X.106.3.551

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55*(4), 627-652. doi: 10.1016/j.jml.2006.08.005

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(5), 489-509. doi: 10.1017/s0140525X08004998

Christiansen, M. H., & Monaghan, P. (2016). Division of labor in vocabulary structure: Insights from corpus analyses. *Topics in Cognitive Science*. *8*(3), 610–624. doi: 10.1111/tops.12164

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. doi: 10.1037/0033-2909.112.1.155

Cornu, G., Mortier, F., Trottier, C., & Bry, X. (2016). SCGLR: Supervised Component Generalized Linear Regression. R package version 2.0.3. https://CRAN.R-project.org/package=SCGLR

de Saussure, F. (1972). *Cours de linguistique générale* (Kobayashi, H., Trans. in Japanese). Tokyo, Japan: Iwanami Shoten. (Original work published in 1916)

Dautriche, I., Mahowald, K., Gibson, E., & Piantadosi, S. T. (2017). Wordform similarity increases with semantic similarity: An analysis of 100 languages. *Cognitive Science, 41*(8), 2149-2169. doi: 10.1111/cogs.12453

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences, 19*(10), 603-615. doi: 10.1016/j.tics.2015.07.013

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review, 119*(2), 223– 271. doi:10.1037/a0027371

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*(6), 627-635. doi: 10.1016/S0022-5371(73)80042-8

Gasser, M. (2004). The origins of arbitrariness of language. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 434–439). Mahwah, New Jersey: Erlbaum.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513-543. doi: 10.1017/S0142716406060383

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(1), 84–95. doi: 10.1037/0278-7393.25.1.84

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences, 23*(5), 389-407. doi: 10.1016/j.tics.2019.02.003

Gupta, P. (2009). A computational model of nonword repetition, immediate serial recall, and nonword learning. In A. Thorn & M. Page (Eds.), Interactions between short-term and long-term memory in the verbal domain (pp. 108–135). Hove, UK: Psychology Press.

Gupta, P., & Tisdale, J. (2009a). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language, 61*(4), 481-502. doi: 10.1016/j.jml.2009.08.001

Gupta, P., & Tisdale, J. (2009b). Word learning, phonological short-term memory, phonotactic probability and long-term memory: Towards an integrated framework. *Philosophical Transactions of the Royal Society: Series B*, *364*(1536), 3755–3771. doi:10.1098/rstb.2009.0132

Henson, R. N. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36*(2), 73-137. doi; 10.1006/cogp.1998.0685

Hockett, C. F. (1960). The Origin of Speech, *Scientific American*, *203*(3), 88–96. doi: 10.1038/scientificamerican0960-88

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term

memory span. *Journal of Memory and Language, 30*(6), 685–701. doi: 10.1016/0749-596X(91)90032-F

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin, 140*(2), 339–373. doi: 10.1037/a0034221

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language, 35*(1), 1-31. doi: 10.1006/jmla.1996.0001

Jefferies, E., Frankish, C., & Noble, K. (2009). Lexical coherence in short-term memory: Strategic reconstruction or "semantic glue"?. *The Quarterly Journal of Experimental Psychology, 62*(10), 1967-1982. doi: 10.1080/17470210802697672

Kelly, M. H. (1992). Using sound to solve syntactic problems: the role of phonology in grammatical category assignments. *Psychological Review*, *99*(2), 349-364. doi: 10.1037/0033-295x.99.2.349

Kindaichi, H. (2001). In Akinaga, K. (Ed.), Shinmeikainihongoakusentojiten [New clear Japanese accent dictionary] (in Japanese). Tokyo, Japan: Sanseido.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*(31) 10681–10686. doi: 10.1073/pnas.0707835105

Kubozono, H. (1995). Gokeiseitoonninkouzou [Word formation and phonological structure] (in Japanese). Tokyo, Japan: Kuroshio shuppan.

Kubozono, H. (2006). *Akusentonohousoku [The rule of accent]* (in Japanese). Tokyo, Japan: Iwanami shoten.

Lambon Ralph, M. A., Moriarty, L., & Sage, K. (2002). Anomia is simply a reflection of semantic and phonological impairments: Evidence from a case-series study. *Aphasiology*, *16*(1-2), 56-82. doi: 10.1080/02687040143000448

Lewandowsky, S., & Murdock, B. B., Jr. (1989). Memory for serial order. *Psychological Review, 96*(1), 25–57. doi: 10.1037/0033-295X.96.1.25

Lewis, M. L., & Frank, M. C. (2016). The length of words reflects their conceptual complexity. *Cognition, 153*, 182-195. doi: 10.1016/j.cognition.2016.04.003

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Attention, Perception, & Psychophysics, 62*(3), 615-625. doi: 10.3758/BF03212113

Majerus, S. (2013). Language repetition and short-term memory: an integrative framework. *Frontiers in Human Neuroscience*, *7*(357). doi: 10.3389/fnhum.2013.0035

Majerus, S., & Boukebza, C. (2013). Short-term memory for serial order supports vocabulary development: New evidence from a novel word learning paradigm. *Journal of Experimental Child Psychology, 116*(4), 811-828. doi: 10.1016/j.jecp.2013.07.014

Majerus, S., Perez, T. M., & Oberauer, K. (2012). Two distinct origins of long-term learning effects in verbal short-term memory. *Journal of Memory and Language, 66*(1), 38-51. doi: 10.1016/j.jml.2011.07.006

Majerus, S., Poncelet, M., Elsen, B., & Van der Linden, M. (2006). Exploring the relationship between new word learning and short-term memory for serial order recall, item recall, and item recognition. *European Journal of Cognitive Psychology, 18*(6), 848-873. doi: 10.1080/09541440500446476

Majerus, S., Poncelet, M., Greffe, C., & Van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology, 93*(2), 95-119. doi: 10.1016/j.jecp.2005.07.005

McQueen, J. M., Otake, T., & Cutler, A. (2001). Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language*, *45*(1), 103–132. doi: 10.1006/jmla.2000.2763

Melby-Lervåg, M., Lervåg, A., Lyster, S. A. H., Klem, M., Hagtvet, B., & Hulme, C. (2012). Nonword-repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychological Science, 23*(10), 1092-1098. doi: 10.1177/0956797612443833

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General, 140*(3), 325-347. doi: 10.1037/a0022924

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130299. doi: 10.1098/rstb.2013.0299

National Institute for Japanese Language and Linguistics. (2015). Balanced Corpus of Contemporary Written Japanese (the short unit list, ersion 1.1). https://ccd.ninjal.ac.jp/bccwj/bcc-chu.html

National Institute for Japanese Language and Linguistics. (2018). UniDic (version 2.3.0). https://unidic.ninjal.ac.jp/download

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for

models of short-term and working memory. *Psychological Bulletin, 144*(9), 885–958. doi: 10.1037/bul0000153

Page, M., & Norris, D. (1998). The primacy model: a new model of immediate serial recall. *Psychological Review, 105*(4), 761-781. doi: 10.1037/0033-295X.105.4.761-781

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1536), 3737-3753. doi: 10.1098/rstb.2009.0173

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526-3529. doi: 10.1073/pnas.1012551108

Reilly, J., & Kean, J. (2007). Formal distinctiveness of high- and low-imageability nouns: Analyses and theoretical implications. *Cognitive Science, 31*(1), 157-168. doi: 10.1080/03640210709336988

Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research*, *47*(1), 223–234. doi:10.1044/1092-4388(2004/019)

Rubin, P., Turvey, M. T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics, 19*(5), 394-398. doi: 10.3758/BF03199398

Saito, S., Nakayama, M., & Tanida, Y. (2020). Verbal working memory, long-term knowledge, and statistical learning. *Current Directions in Psychological Science*, *29*(4), 340-345. doi: 10.1177/0963721420920383

Sakono, S., Ito, T., Fukuda, S. E., & Fukuda, S. (2011). The effect of word accent production on reading performance in Japanese young children. *Asia Pacific Journal of Speech, Language, and Hearing*, *14*(1), 51–59. doi: 10.1179/136132811805334948

Sakuma, N., Ijuin, M., Fushimi, T., Tatsumi, I., Tanaka, M., Amano, S., & Kondo, T., (2005). *NTT database series: Lexical properties of Japanese, Vol. 8* (in Japanese). Tokyo, Japan: Sanseido.

Schmidtke, D., Conrad, M., & Jacobs, A. M. (2014). Phonological iconicity. *Frontiers in Psychology, 5*, 80. doi: 10.3389/fpsyg.2014.00080

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, *21*(2), 168–175. doi:10.3758/BF03202729

Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences, 14*(9), 411–417. doi: 10.1016/j.tics.2010.06.006

Smith, K., Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2008). Introduction. Cultural transmission and the evolution of human behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1509), 3469–3476. doi: 10.1098/rstb.2008.0147

Tanida, Y., Nakayama, M., & Saito, S. (2019). The interaction between temporal grouping and phonotactic chunking in short-term serial order memory for novel verbal sequences. *Memory, 27*(4), 507-518. doi: 10.1080/09658211.2018.1532008

Tanida, Y., Ueno, T., Ralph, M. A. L., & Saito, S. (2015a). The roles of long-term phonotactic and lexical prosodic knowledge in phonological short-term memory. *Memory & Cognition*, *43*(3), 500-519. doi: 10.3758/s13421-014-0482-2

Tanida, Y., Ueno, T., Lambon Ralph, M. A., & Saito, S. (2015b). The influence of accent pattern typicality on immediate and delayed nonword repetition. *Psychologia*, *58*(3), 145-154. doi: 10.2117/psysoc.2015.145

Thorn, A. S. C., Gathercole, S. E., & Frankish, C. R. (2005). Redintegration and the benefits of long-term knowledge in verbal short-term memory: An evaluation of Schweickert's (1993) multinomial processing tree model. *Cognitive Psychology, 50*(2), 133–158. doi: 10.1016/j.cogpsych.2004.07.001

Thorn, A., & Page, M. (2008). Interactions between short-term and long-term memory in the verbal domain. Psychology Press.

Tomaschek, F., Hendrix, P., & Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics, 71*, 249-267. doi: 10.1016/j.wocn.2018.09.004

Tyler, L. K., Voice, J. K., & Moss, H. E. (2000). The interaction of meaning and sound in spoken word recognition. *Psychonomic Bulletin & Review, 7*(2), 320–326. doi: 10.3758/bf03212988

Ueno, T. (2012). The impact of conceptual knowledge on phonological processing in Japanese. Unpublished doctoral dissertation, The University of Manchester, UK. Retrieved from www.escholar. manchester.ac.uk/uk-ac-man-scw:157460

Ueno, T., Saito, S., Rogers, T. T., & Lambon Ralph, M. A. (2011). Lichtheim 2: synthesising aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, *72*(2). 385–96. doi: 10.1016/j.neuron.2011.09.013

Ueno, T., Saito, S., Saito, A., Tanida, Y., Patterson, K., & Lambon Ralph, M. A. (2014). Not lost in translation: Generalization of the primary systems hypothesis to

Japanese-specific language processes. *Journal of Cognitive Neuroscience*, *26*(2), 433–446. doi:10.1162/jocn_a_00467

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374-408. doi: 10.1006/jmla.1998.2618

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, *40*(1), 47–62. doi: 10.1177/002383099704000103

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge, England: Addison-Wesley.

Tables

Table 1. Accent types for each word length

	Accent type					
	Flat	Type 1	Type 2	Type 3	Type 4	…
1-mora	L-(H) 64	H-(L) 290				
2-mora	L-H-(H) 230	H-L-(L) 764	L-H-(L) 198			
3-mora	L-H-H-(H) 2,978	H-L-L-(L) 2,340	L-H-L-(L) 171	L-H-H-(L) 254		
4-mora	L-H-H-H-(H) 6,887	H-L-L-L-(L) 1,031	L-H-L-L-(L) 829	L-H-H-L-(L) 375	L-H-H-H-(L) 30	
⋮						

Note: In the table, ‘H’ and ‘L’ represent high- and low-pitch mora, respectively. The final L or H in parentheses represents the pitch of the postpositional particle following the word. Postpositional particles are composed of one mora in most cases, e.g., ‘ga’ and ‘wa’ representing nominatives or ‘wo’ representing objectives. For tri-mora words, for example, the flat accent does not have the pitch drop, but type 3 does at the end of the final mora. Values in each cell represent the number of common noun entries with each accent type of each length. Entries were counted in our 18,262-noun dataset for 2- to 4-mora nouns but in all 101,603 common nouns in BCCWJ for 1-mora nouns.

Table 2. Abbreviations for variables

Label		Variable
acr		accent regularity
	flt	flat accent or not
	reg	regularly accented pattern or not
bmf		bi-mora frequency
img		imageability
len		word length (number of phonemes)
nbh		neighbourhood size
smf		single-mora frequency
wdf		word frequency
wdt		etymological word type
	njp	factor of native Japanese word or not
	sjp	factor of Sino-Japanese word or not
	lnw	factor of loanword or not
	hyb	factor of hybrid word or not

Table 3. Outcomes of Brunner-Munzel test and effect size (Cliff's delta) for each accent regularity category included in each dataset

		Regularly accented pattern	Irregularly accented pattern
The whole dataset	Flat	W = 14.027, df = 10454, p < .001, *Cliff's delta* = -0.134 [-0.152, -0.115]	W = 16.828, df = 2056.3, p < .001, *Cliff's delta* = -0.259 [-0.289, -0.229]
	Regularly accented pattern		W = 7.002, df = 2835.1, p < .001, *Cliff's delta* = -0.108 [-0.139, -0.078]
Without loanwords	Flat	W = 0.148, df = 6522.8, p = .88, *Cliff's delta* = -0.002 [-0.023, 0.020]	W = 13.348, df = 1287.5, p < .001, *Cliff's delta* = -0.249 [-0.285, -0.212]
	Regularly accented pattern		W = 12.164, df = 1774.7, p < .001, *Cliff's delta* = -0.230 [-0.266, -0.192]

Note. Values in brackets represent the 95% confidence interval of the effect size.

Table 4. Result of the ordinary least regression analyses and the back-translated coefficient from the SCGLR morel.

		ordinary least regression with all variables					ordinary least regression without neighbourhood size							SCGLR			
		beta	*SE*	*t*	*p*	*VIF*	*beta*	*SE*	*t*	*p*	*VIF*	Likelihood ratio test *F*	Likelihood ratio test *p*	*back-translated beta*	*SE*	*t*	*p*
The whole dataset	intercept	-0.507	0.025	-20.398			-0.486	0.024	-19.832					-0.480			
	bmf	-0.048	0.009	-5.272	< .001	2.099	-0.055	0.009	-6.020	< .001	2.057	36.246	< .001	-0.008			
	len	-0.089	0.013	-6.667	< .001	4.518	-0.036	0.008	-4.556	< .001	1.604			-0.011			
	nbh	-0.078	0.016	-4.919	< .001	6.381			—					0.005			
	smf	-0.011	0.008	-1.295	0.195	1.681	-0.011	0.008	-1.374	0.169	1.681			-0.052			
	wdf	0.345	0.007	50.372	< .001	1.174	0.343	0.007	50.137	< .001	1.170			0.341			
	acr: flt	0.120	0.026	4.674	< .001	1.452	0.075	0.024	3.120	0.002	1.266			0.068	0.024	2.822	0.005
	acr: regular	0.079	0.025	3.133	0.002		0.041	0.024	1.720	0.086				0.034	0.024	1.418	0.156
	wdt: njp	0.834	0.021	39.433	< .001		0.873	0.020	44.323	< .001				0.873	0.019	46.292	< .001
	wdt: lnw	1.053	0.022	46.821	< .001	2.554	1.086	0.021	50.561	< .001	2.114			1.087	0.022	49.461	< .001
	wdt: hyb	0.677	0.035	19.445	< .001		0.724	0.033	21.665	< .001				0.730	0.034	21.752	< .001
		Residual SE = 0.853, adjusted R2 = 0.272, F(10, 18251) = 682.877, p < .001					Residual SE = 0.854, adjusted R2 = 0.271, F(9, 182512) = 755.105, p < .001										

Without loanwords	intercept	-0.247	0.030	-8.340			-0.240	0.029	-8.202					-0.303			
	bmf	-0.102	0.011	-9.479	< .001	2.409	-0.108	0.010	-10.586	< .001	2.143	112.058	< .001	-0.022			
	len	0.016	0.012	1.372	0.170	2.854	0.029	0.008	3.587	< .001	1.358			-0.126			
	nbh	-0.025	0.016	-1.522	0.128	5.583			—					0.113			
	smf	0.004	0.009	0.432	0.666	1.686	0.005	0.009	0.569	0.569	1.673			-0.065			
	wdf	0.405	0.008	52.604	< .001	1.228	0.405	0.008	52.600	< .001	1.225			0.191			
	acr: flt	-0.023	0.031	-0.739	0.460	1.293	-0.040	0.029	-1.404	0.160	1.098			-0.065	0.029	-2.212	0.027
	acr: regular	-0.073	0.031	-2.347	0.019		-0.085	0.030	-2.840	0.005				-0.114	0.031	-3.663	< .001
	wdt: njp	0.802	0.024	33.174	< .001	2.639	0.823	0.020	41.973	< .001	1.698			1.079	0.020	53.139	< .001
	wdt: hyb	0.683	0.036	18.768	< .001		0.704	0.034	20.765	< .001				0.860	0.036	24.170	< .001
		Residual SE = 0.862, *adjusted* R^2 = 0.257, $F(9, 15355) = 590.235$, $p < .001$					*Residual SE* = 0.862, *adjusted* R^2 = 0.257, $F(8, 15356) = 663.668$, $p < .001$										

Figures

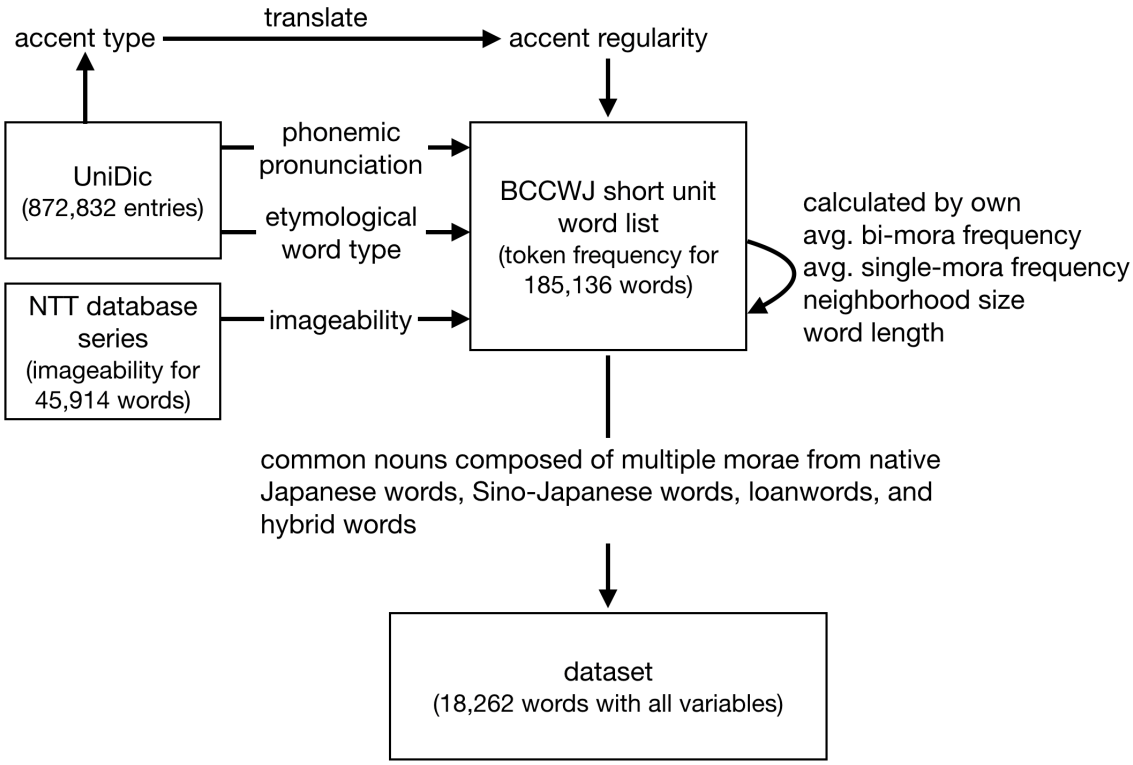
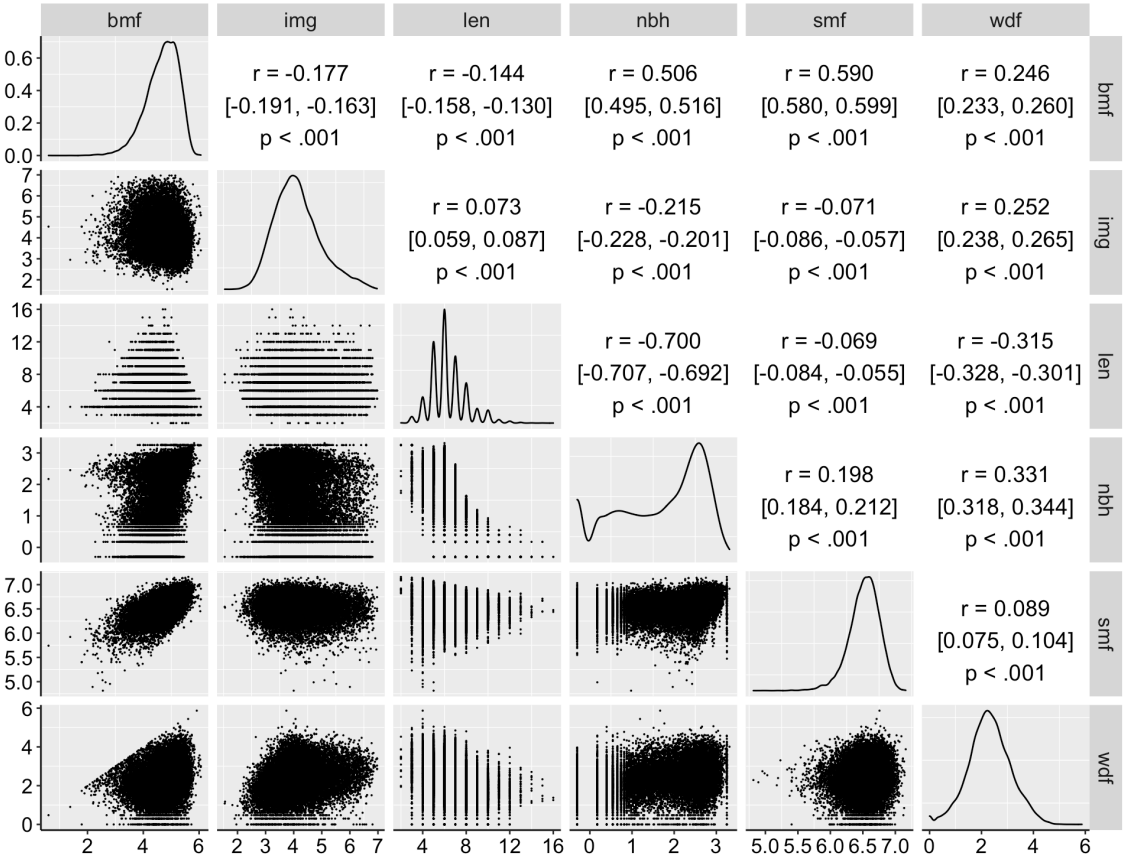


Figure 1. Schematic image of dataset preparation

The whole dataset



Without loanword

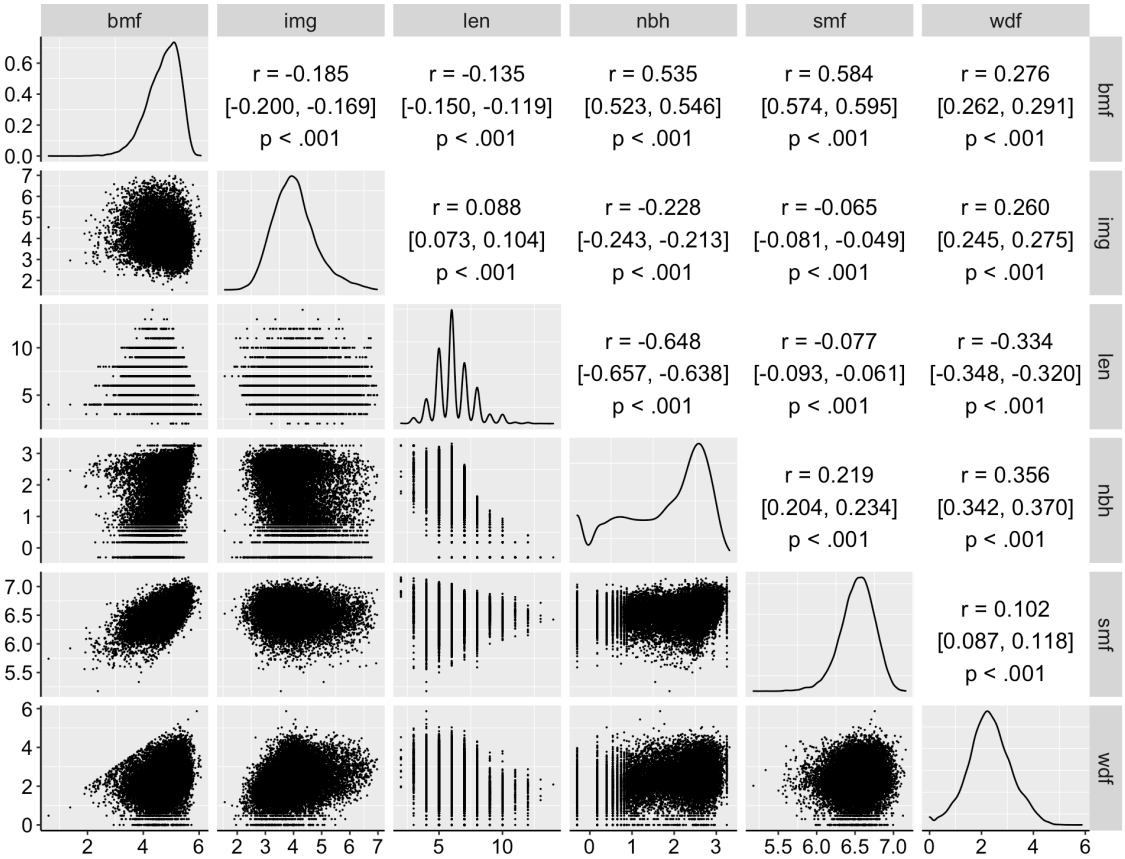


Figure 2. Correlation plots of continuous variables in the whole dataset and the dataset without loanwords. A plot represents a noun. The results of Spearman's rank correlation analyses are provided. Values in brackets represent 95% confidence intervals for *r*.

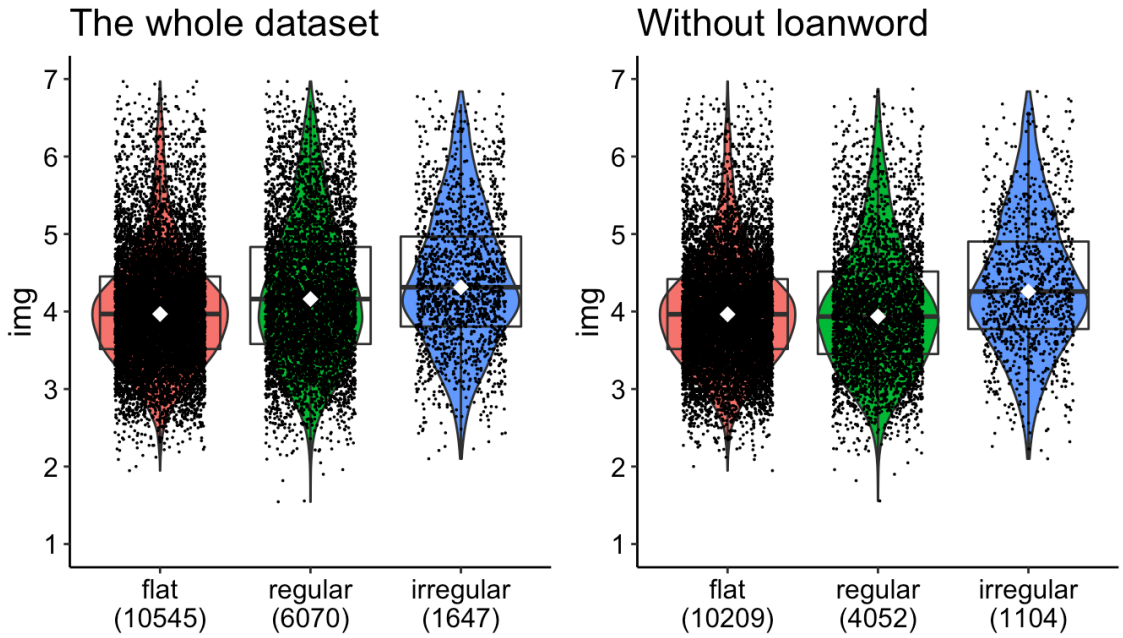


Figure 3. Box- and violin-plots for median imageability for each accent category in the whole dataset and the dataset without loanwords. A plot represents a noun. Values shown under levels are type frequencies for each level.

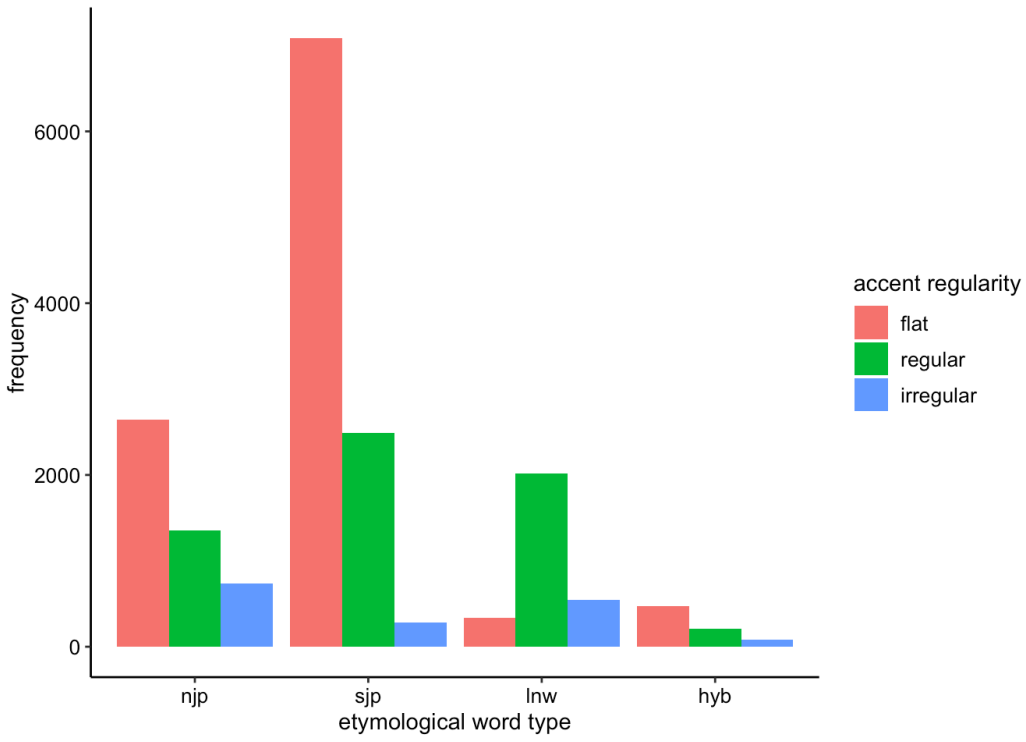


Figure 4. Type frequency of each accent pattern in each etymological word type

Supplementary Information

A. Detailed explanation and procedure of the SCGLR analysis

Regression analysis with multiple variables can suffer from multicollinearity (Tomaschek et al., 2018; Westfall & Yarkoni, 2016). A critical problem in regression analysis controlling collinear multiple variables is the suppression and enhancement of the regression coefficient (beta). Tomaschek et al. (2018) presented a representative example whereby reaction time in auditory lexical decisions is predicted by two types of word frequency counted in American English and British English. Reaction time negatively correlates with both types of word frequency. However, because of the strong positive correlation between American and British word frequency, a regression analysis with these two variables can give the expected negative coefficient to American English frequency but a positive coefficient to British English frequency.

To mitigate such collinearity problems, Tomaschek et al. (2018) introduced a coefficient shrinkage method, *SCGLR analysis*, and the R package *SCGLR* (Cornu et al., 2016). SCGLR is conceptually based on principal component regression (PCR), which relies on principal component analysis. Principal component analysis successively abstracts orthogonal principal components capturing the largest variances in the original variables. PCR is used when a dependent variable is predicted from a set of variables with collinearity. In PCR, principal components taken from the set of original variables by principal component analysis are projected for independent variables in a regression model predicting the dependent variable. Imagine that you originally have N independent variables, and a principal component analysis of the original N variables obtains the N principal components. Then, when the N principal components are projected into a regression model, the predictive power of the PCR model for the dependent variable is equivalent to that of the ordinary least square

regression model with the original N variables. This is the case because variance across the N components keeps variance across the original N variables. Importantly, collinearity does not occur in the PCR model because all principal components used for predictors in the model are orthogonal. Another important characteristic of PCR is that the number of components to be projected into the regression model can be reduced. For example, when a principal component analysis shows that a certain level (e.g., 90%) of variance is explained by principal components 1 to X among N variables (here, X < N), the PCR model with the X components for predictors predicts the dependent variable from the reduced variance while sacrificing predictive power. Consequently, the model reduces overfitting deriving from collinearity across the original independent variables.

The difference between PCR and SCGLR lies in the range of orthogonalization. PCR orthogonalizes only the original independent variables and does not include the dependent variable. In contrast, SCGLR orthogonalizes the original independent variables and dependent variable collectively in building the regression model. That is, components are obtained while considering the variance of the dependent variable. This is why components in SCGLR are called *supervised* components (abbreviated as *sc* here) instead of principal components. Additionally, the SCGLR package implements a cross-validation procedure to determine the optimal number of components, although the number of employed components used in the conventional principal component analysis is determined by the rule of thumb. Although an SCGLR analysis estimates regression coefficients for each supervised component of predictors, one can translate the supervised component coefficients into the coefficients for each original independent variable, which are called back-translated coefficients. These back-translated coefficients can be recognized as coefficients for the original independent variable after managing problems of collinearity.

To obtain these back-translated coefficients for the original independent variables, we performed an SCGLR analysis with a Gaussian distribution for family functions on our dataset following the procedure introduced in Tomaschek et al. (2018). All SCGLR analyses presented in this paper were performed with the SCGLR package (version 2.0.3, Cornu et al., 2016) with R (version 3.6.3, R core team, 2020) and RStudio (version 1.4.1106, RStudio Team, 2021) in macOS Catalina (version 10.15.7). Here, we provide an overview of the SCGLR analysis conducted with our whole dataset, including loanwords. Variables to be analysed were prepared in the same way as in our OLS regression analyses described in the main text. The dependent variable for the SCGLR model was standardized imageability. All five continuous variables, including bi-mora frequency, single-mora frequency, word frequency, word length (the number of phonemes), and neighbourhood size, were standardized. Two categorical variables, accent regularity and etymological word type, were coded by dummy coding such that the intercept was the level of irregularly accented patterns and Sino-Japanese words.

Importantly, only continuous variables can be orthogonalized when building a model with supervised components as predictors in SCGLR analysis, and categorical variables cannot, as in conventional principal component analysis, in which binary variables cannot be orthogonalized jointly with other variables. Therefore, all continuous variables were subjected to orthogonalization in the SCGLR model, but accent regularity and etymological word type were out of orthogonalization and projected for covariates to the model. To determine the optimal number of supervised components (similarly to determining the number of principal components considered in principal component analysis), we performed cross-validation through the function *scglrCrossVal* in the SCGLR package. The maximum number of components to take

into account was five (the number of predictors to be orthogonalized). The number of folds was set to 10. We performed the same cross-validation procedure 11 times while varying random seeds. Then, we averaged the optimal number of components across the 11-times cross-validation and rounded the averaged value. Finally, we employed supervised components 1 to 3. At this stage, we obtained (1) the relationship between abstracted supervised components and the original independent variables presented in Table A1 and Figure A1 (corresponding to principal component loading and the principal component score in principal component analysis) and (2) the estimated regression model predicting the dependent variable (imageability) from the three supervised components and the categorical variables, as presented in Table A2.

Table A1 presents information on the relationships between the original variables and the supervised components employed. The squared correlations (r^2) of the original independent variables with each supervised component show that the bi-mora frequency and single-mora frequency were most represented on the plane composed of supervised components 1 and 2, as listed in the column of the best plane in Table A1. Imageability, word length, neighbourhood size, and word frequency were most represented on the plane composed of supervised components 1 and 3. Thus, supervised component 2 strongly represents phonotactic regularity relative to supervised component 3. The best value in Table A1 indicates that variances for all predictors were sufficiently explained by their best planes (0.717~0.934). In the correlation plot of Figure A1, arrows represent the values and signs of correlations (r) between the original independent variables and supervised components. We can see that on the axis of supervised component 2, the bi-mora frequency shows a negative value, but imageability shows a positive value, indicating that the bi-mora frequency is negatively correlated with imageability in supervised component 2. On the other hand, bi-mora

frequency and imageability were positively correlated in supervised component 1, and their correlation was nearly zero in supervised component 3.

Table A1. Squared correlations between the original variables and supervised components

		sc1	sc2	sc3	Best plane	Best value
Whole dataset	img	0.389	0.161	0.450	1/3	0.839
	bmf	0.361	0.465	0.001	1/2	0.826
	len	0.543	0.163	0.174	1/3	0.717
	nbh	0.720	0.022	0.147	1/3	0.867
	smf	0.151	0.662	< 0.001	1/2	0.812
	wdf	0.447	0.063	0.488	1/3	0.934
Without loanwords	img	0.751	0.249		1/2	1.000
	bmf	0.337	0.485		1/2	0.822
	len	0.520	0.109		1/2	0.629
	nbh	0.691	0.002		1/2	0.693
	smf	0.127	0.674		1/2	0.800
	wdt	0.558	0.077		1/2	0.634

Note. The columns for sc1, sc2, and sc3 show the squared correlations (r^2) of the original predictors with each supervised component. The column of best plane presents two supervised components in which the variance of each variable was best explained. The column of best value represents the sum of the r^2 values for the axes of the best plane. The value indicates the variance in the response captured by the best plane.

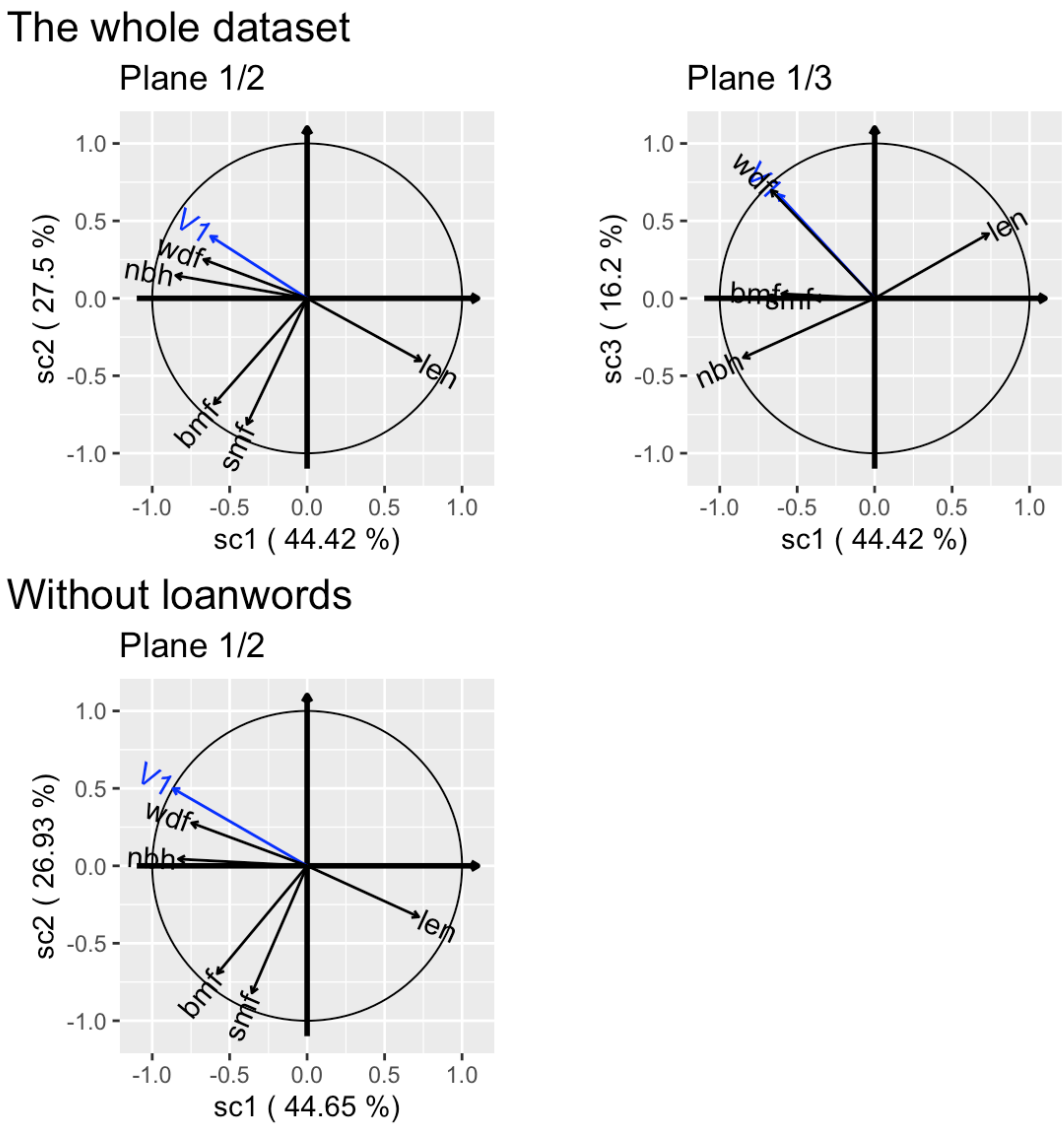


Figure A1. Correlation plots for variables on the planes composed of three supervised components in analyses for each dataset. The dependent variable (imageability) is labelled ‘V1’ in blue font.

The estimated SCGLR model predicting the dependent variable from the three supervised components and the two categorical variables is shown in Table A2. As a result of the conventional multiple regression analysis, we obtained the controlled regression coefficients and their significance for each supervised component packaging continuous variable and for each term deriving from the categorical variables.

Therefore, we did not obtain the controlled coefficients and their significance for the *original* independent variables but obtained them for categorical variables at this stage. For etymological word type, the model estimated that Sino-Japanese words (the level of the intercept) showed significantly lower imageability than all other word types (*p*s <.05). In contrast, we obtained the controlled coefficients for accent regularity against our prediction and the results of the Brunner-Munzel tests (Table 3). The model indicates that the imageability of irregularly accented nouns was significantly higher than that of flat nouns. The imaginability of irregularly accented nouns was also higher than that of regularly accented nouns, although this coefficient was not significant.

Table A2. Coefficient of supervised components and categorical variables in the SCGLR model and the inertia for each supervised component

		beta	*SE*	*t*	*p*	Inertia
Whole dataset	intercept	-0.480	0.024	-19.811	< .001	
	sc1	-0.146	0.005	-26.967	< .001	0.444
	sc2	0.118	0.005	21.796	< .001	0.275
	sc3	0.257	0.007	34.993	< .001	0.162
	acr: flt	0.068	0.024	2.822	.005	
	acr: regular	0.034	0.024	1.418	.156	
	wdt: njp	0.873	0.019	46.292	< .001	
	wdt: lnw	1.087	0.022	49.461	< .001	
	wdt: hyb	0.730	0.034	21.752	< .001	
Without loanwords	intercept	-0.303	0.030	-10.104	< .001	
	sc1	-0.199	0.006	-31.690	< .001	0.446
	sc2	0.145	0.006	23.100	< .001	0.269
	acr: flt	-0.065	0.029	-2.212	.027	
	acr: regular	-0.114	0.031	-3.663	< .001	
	wdt: njp	1.079	0.020	53.139	< .001	
	wdt: hyb	0.860	0.036	24.170	< .001	

Next, we examine the coefficients for three supervised components involving information for the original continuous variables. The model shows that the dependent variable of imageability was significantly predicted by each supervised component (*p*s <.05 in Table A2). Importantly, SCGLR can calculate the regression coefficients for *the original independent variables* by back-translation from the regression coefficients for the supervised components (Table A2) via correlations between each supervised component and original variable (Table A1). The back-translated coefficient for the original independent variables is moderated by dimensionality reduction on the variables' space when we discard two supervised components of the maximum of five components (that is, when employing only three supervised components). Therefore, the back-translated coefficients for the original independent variables can be recognized as shrunken/regularized coefficients, that is, coefficients for which enhancement and suppression were inhibited as much as possible. The back-translated coefficients for the original independent variables are shown in Table 4 in the main text, where the coefficients for categorical variables were not back-translated but retrieved from Table A2.

Finally, while outside the scope of SCGLR analysis but for confirmation, we examined the linearity of the significant-negative relationship between imageability and bi-mora frequency. It is argued that even if one obtains a significant result from correlation analysis and/or linear regression, the relationship of two variables may be nonlinear, such as L- or inverted L-shaped. The whole dataset was divided into four datasets for each quartile of the bi-mora frequency variable. Figure A2 shows violin plots and the median and interquartile range of imageability in the first- to fourth-quartile datasets. We can see that the median imageability was the highest in the lowest bi-mora frequency group (the dataset of the first quartile of bi-mora frequency), and the

median imageability decreased as the bi-mora frequency increased. Therefore, imageability linearly increases as bi-mora frequency decreases in Japanese vocabulary.

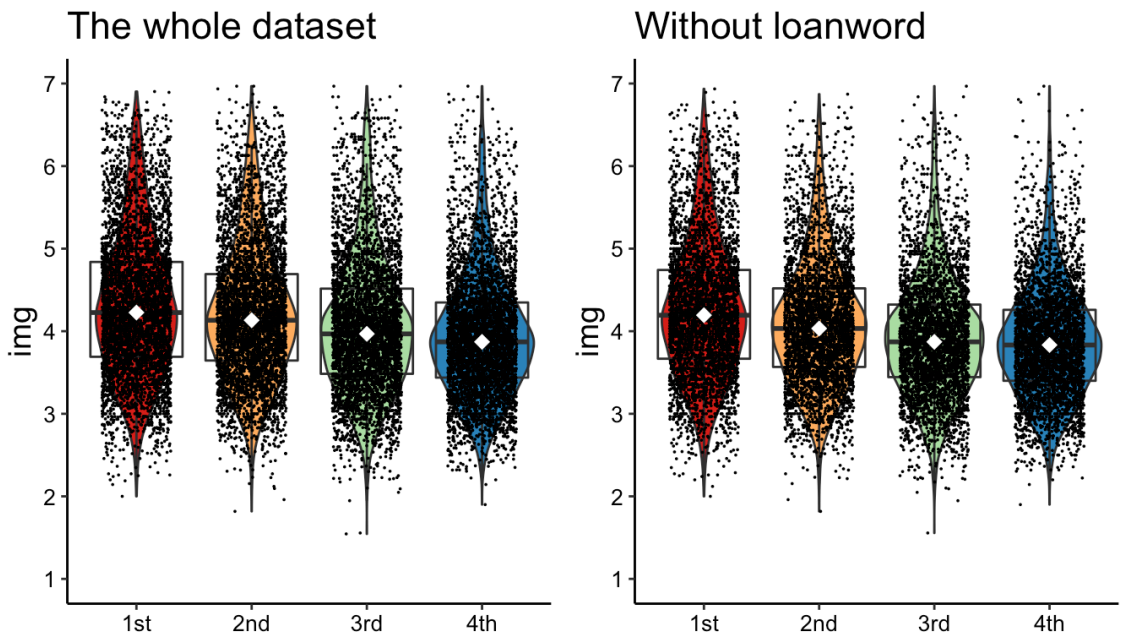


Figure A2. Box-and violin-plot for median imageability of nouns whose bi-mora frequency belonged to each quartile in each dataset. The white squares represent medians.

Analyses of the dataset without loanwords were performed in the same way as those for the whole dataset, except that we employed, for the SCGLR model, two supervised components according to the result of cross-validation. The results are presented in Tables A1 and A2 and Figures A1 and A2 in the same way as the results of the whole dataset.

B. A potential factor fuzzifying the predicted negative correlation between imageability and accent regularity

A factor that may fuzzify the predicted negative correlation between imageability and accent regularity is the influence of *accenting* an element in a memory sequence. Verbal working memory research has shown that accenting an element of a sequence affects short-term *phonemic* representation, interacting with the serial position effect. An example is provided by Gupta, Lipinski, Abbs, and Lin (2005), showing the serial position effect that occurs when repeating a multisyllabic English nonword. The authors report that the repetition accuracy for a stressed syllable at middle serial positions was specifically higher than that for unstressed neighbouring syllables. Consequently, the serial position curve showed a W-shaped curve with the three tops at the typical primary and recency positions and the accented position. This accenting effect on the accented syllable was obtained even in English-speaking children (Chiat & Roy, 2007; Roy & Chiat, 2004), although the W-shaped serial position curve was not observed because the authors employed fewer than 4-syllabe nonwords. Reeves, Schmauder, and Morris (2000) found this accenting effect in immediate serial recall. The authors compared the recall accuracy of 9-item lists under two conditions. In the uniform-stress condition, all items in a list were stressed uniformly. In the stressed condition, the first, fourth, and seventh items or third, sixth, and ninth items were stressed. Recall performance was higher for items at the stressed positions, especially those located at middle positions, than items at the same serial positions in the uniform-stress condition. The accenting effect was found even in Japanese, which is a pitch accent language. Yuzawa and Saito (2006) found accenting a mora to improve the repetition performance of the accented mora in 2-mora nonwords in children.

This accenting effect can be a means to overcome the vulnerability deriving from the serial position effect. It is well known that recall performance is lower at middle positions than at other positions in a list/sequence for verbal working memory tasks (e.g., Botvinick & Plaut, 2006; Gupta, 2005; Gupta et al., 2005; Hurlstone et al., 2014; Ward, Tan, & Grenfell-Essen, 2010). The decrease in available information at the middle positions is critical for learning and communicating whole word sequences and therefore for survival through cultural transmission. It is noteworthy that the probability of holding the whole word form can be improved by accenting syllables, particularly those at middle positions. Thus, it might be possible that the linguistic environment reflects the effect of accenting to maintain the availability of phonemic representation in addition to the effect of accent regularity on the quality of accent representation through accent regularity knowledge. To investigate this potential factor, we counted the type frequency of nouns in our two datasets as a function of word length (the number of syllables) and accented syllabic position (as counted from the onset). Figure B1 shows that regardless of the datasets, flat nouns are the most frequent in general. The point we would like to emphasize is that for longer words, the flat accent pattern is not dominant. Furthermore, among accented nouns, the accented position shifts from the first syllable in shorter words to the middle-position syllables in longer words, shown as salient inverse U-shaped curves in four- to six-syllable accented words. Under such a lexical environment, the vulnerability of phonemic information at middle positions, particularly in longer words, should be compensated for by the accenting effect, leading to the successful retention of the whole sequence in verbal working memory.

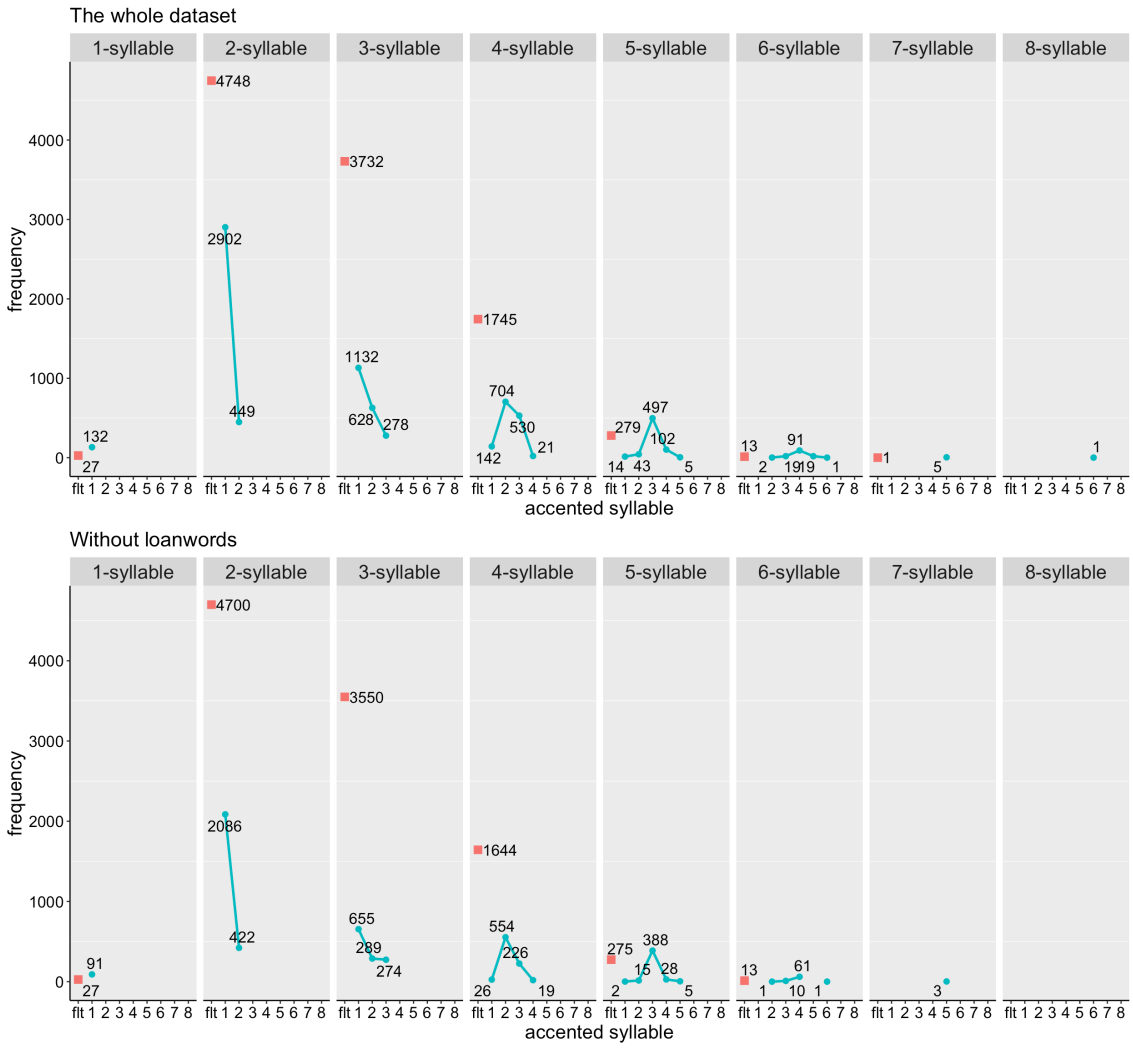


Figure B1. Frequency of common nouns in each dataset as a function of word length and accented syllable positioning counted from the onset. The values represent the type frequency for each plot. Red square plots represent flat nouns, and green circle plots represent accented nouns.

Remarkably, for words with four or more syllables, the most dominant accent position departs from the first syllable. This boundary is compatible with evidence that the significant serial position effect appears for a sequence with four syllables or more in nonword repetition tasks (Gupta et al., 2005) and in a sequence with four digits or more in immediate serial recall tasks (Ward et al., 2010). Furthermore, the fact that the serial position effect appears from a 4-item length is consistent with more recent

theories of the limit of working memory capacity. Miller's classic magical number seven plus/minus two (Miller, 1957) is known as the maximum limit of immediate memory when participants use a sort of strategy, such as chunking and rehearsal. Cowan (2001, ass also, Chen & Cowan, 2005; 2009) predicted that capacity is more limited to approximately four without such strategies. Therefore, the Japanese vocabulary space can be referred to as an example case in which words of the maximum length beyond the restricted working memory capacity of four items without strategies can survive by a "natural" strategy of accenting a syllable at middle positions.

This hypothesis of the accenting effect in the lexical environment is applicable to the English accentual structure. A corpus analysis study by Reilly and Kean (2007) reports that in English, accent patterns for shorter and highly imageable words tend to be those with a fixed stress position at the typical within word position (i.e., the first syllable), whereas longer or low-imageable words show more variations with less systematicity in their stress position. For two-syllable words, their stress was found to be located at the first syllables for 93.5% of high-imageable words and 75.3% of low-imageable words. Three-syllable words more strongly showed this tendency of stronger stress position dispersion for low-imageable words. Among high-imageable words, 72.78% of words stressed the first syllable, 22.78% stressed the second syllable, and 4.44% stressed the third syllable. Among low-imageable words, 47.80%, 50.00%, and 2.20% of words stressed at the first, second, and third syllables, respectively. Additionally, both high- and low-imageable words stressed various syllables in four-syllable words. The first, second, third, and fourth syllables were stressed for 34.23%, 42.86%, 20.00%, and 2.86% of the high-imageable words, respectively, and for 34.23%, 45.05%, 18.92%, and 1.81% of the low-imageable words, respectively. These accentual structures in English words are not useful for preserving the quality of short-

term *accent* representation through the compensatory contributions of semantic and accent regularity knowledge. Rather, the English accentual structure might have a mechanism that can increase the availability of short-term *phonemic* representations at middle positions in low-imageable longer words. In sum, both the Japanese and English accent structures might depend on effects related to verbal working memory properties but in different ways.

References

Chen, Z., & Cowan, N. (2005). Chunk limits and length limits in immediate recall: a reconciliation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(6), 1235-1249. doi: 10.1037/0278-7393.31.6.1235

Chen, Z., & Cowan, N. (2009). Core verbal working-memory capacity: The limit in words retained without covert articulation. *Quarterly Journal of Experimental Psychology, 62*(7), 1420-1429. doi: 10.1080/17470210802453977

Chiat, S., & Roy, P. (2007). The Preschool Repetition Test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language & Hearing Research, 50*(2), 429-443. doi: 10.1044/1092-4388(2007/030)

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and brain sciences, 24(1), 87-114. dot: 10.1017/S0140525X01003922

Gupta, P. (2005). Primacy and recency in nonword repetition. *Memory, 13*(3-4), 318-324. doi: 10.1080/09658210344000350

Gupta, P., Lipinski, J., Abbs, B., & Lin, P. H. (2005). Serial position effects in nonword repetition. *Journal of Memory and Language, 53*(1), 141-162. doi: 10.1016/j.jml.2004.12.002

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97. doi: 10.1037/h0043158

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reeves, C., Schmauder, A. R., & Morris, R. K. (2000). Stress grouping improves performance on an immediate serial list recall task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1638-1654. doi: 10.1037/0278-7393.26.6.1638

RStudio Team (2021). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Ward, G., Tan, L., & Grenfell-Essam, R. (2010). Examining the relationship between free recall and immediate serial recall: the effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1207-1241. doi: 10.1037/a0020122

Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLOS ONE, 11*(3), e0152719. doi: 10.1371/journal.pone.0152719

Yuzawa, M., & Saito, S. (2006). The role of prosody and long-term phonological knowledge in Japanese children's nonword repetition performance. *Cognitive Development, 21*(2), 146-157. doi: 10.1016/j.cogdev.2006.01.003