

ORIGINAL ARTICLE

Phonological processing of Japanese *Kanji* word by L1-Chinese learners of Japanese: verification of consistency and frequency effects

Nan Kang  and Satoru Saito 

Graduate School of Education, Division of Cognitive Psychology in Education, Kyoto University, Kyoto, Japan

Corresponding author: Nan Kang; Email: kangnan70@gmail.com

(Received 29 January 2024; revised 20 January 2025; accepted 22 March 2025)

Abstract

Studies investigating phonological processing indicate that words with high regularity/consistency in pronunciation or high frequency positively impact reading speed and accuracy. Such effects of consistency and frequency have been demonstrated in Japanese kanji words and are known as consistency and frequency effects. Using a mixed-effects model analysis, this study reexamines the two effects in Chinese–Japanese second-language (L2) learners with two different L2 proficiency levels. The two effects are robustly replicated in oral reading tasks; in particular, the performance of intermediate learners is similar to that of Japanese semantic dementia patients, whose reading accuracy is affected by sensitivity to the statistical properties of words (i.e., reading consistency and lexical frequency). These results are explained by the interaction between semantic memory and word statistical properties. Moreover, the interaction highlights the important consequences of statistical learning underlying L2 phonological processing.

Keywords: Japanese kanji word; L2 learning; lexical frequency; reading consistency; statistical learning

Introduction

The phonological processing of logographic writing systems has been receiving scholarly attention for decades. One exemplification is kanji words in the Japanese writing system, which comprises syllabic kana and kanji. Unlike the translation rules for print-to-sound in alphabetic writing systems (e.g., English), the orthography-to-phonology (O-P) correspondence of Japanese kanji words is less regular and systematic. For instance, in the phrase 地球は丸い (the earth is round), the subject 地球 and the first character of the adjective 丸い are written in kanji, whereas the remaining are in kana script. The morphology of kana (e.g., い /i/) maps to its phonological representation with one syllabic unit, *mora* (Otake et al., 1993).

However, a single Japanese kanji can have more than one legitimate pronunciation due to its intra-context within the word and, hence, cannot be decomposed to match any specific size of a syllabic unit. Thus, whether a set of rules for O-P mapping regularity can be applied to the reading of Japanese kanji words is unclear.

Reading of Japanese Kanji words

Studies have addressed O-P mapping regularity in terms of *on*- and *kun*-reading in Japanese kanji (Ida et al., 2014; Wydell et al., 1995). *On*-reading, where a word's pronunciation can be read from its character level, derives from the pronunciation of original Chinese characters. *Kun*-reading, where the constituent character's pronunciation obeys less corresponding regularity, derives from the pronunciation of an original Japanese word. For instance, 地球 /chi-kyu/ is an *on*-reading word, whereas 丸 /maru/ is a *kun*-reading character in the word 丸 子. Additionally, kanji words like 地球 /chi-kyu/ can be read from their sub-word level. That is, the constituent characters 地 and 球 are considered to map the pronunciations /chi/ and /kyu/, respectively, as read in the words 地形 /chi-kei/ (topography) and 円球 /eN-kyu/ (sphere). Compared with *kun*-reading, *on*-reading is a more regular type of pronunciation (Fushimi et al., 1999; Wydell et al., 1995). Although the *on*-reading of kanji was generally preserved during its introduction from ancient China, many words lost their original Chinese pronunciations in terms of accents or tones (Wydell et al., 1993). The pronunciation of Japanese kanji itself became complex such that one character could have several legitimate, alternative *on*-reading pronunciations. For instance, the character 地 is also pronounced as /ji/ in the word 地面 /ji-meN/ (the ground). Hence, *on*-reading cannot prescribe the regularity of O-P correspondences in Japanese kanji words (Fushimi et al., 1999; Ida et al., 2014; Patterson et al., 1995). To allocate the appropriate pronunciation when the constituent kanji character has multiple legitimate pronunciations, one must harness phonological knowledge at the whole-word level where contextual and lexical-semantic knowledge is supposed to be activated (Fushimi et al., 1999; Fushimi et al., 2009; Wydell et al., 1993).¹

Reading consistency in Japanese Kanji words

Another perspective to capture the regularity of the O-P correspondence in kanji words is the consistency/typicality of the pronunciation. It is described as *reading consistency* and has been discussed mainly in the context of two-character Japanese kanji words (Fushimi et al., 1999; Wydell et al., 1993; Wydell et al. 1995). Fushimi et al. (1999) defined reading consistency by computing the size of the phonological and orthographical neighborhood (characters having identical phonological/orthographical forms, respectively) of the constituent kanji character within the target word. By statistically analyzing a corpus of approximately 31,000 words from the *Iwanami Japanese Dictionary* (Nishio et al., 1986), the study categorized reading consistency into consistent, inconsistent-typical, and inconsistent-atypical words.² When the constituent kanji has identical pronunciation among all the neighbors including the same character in the same position, it is considered a consistent word. By contrast, when the constituent kanji has more than one

legitimate pronunciation, it is an inconsistent word. Inconsistent words are further classified as typical and atypical words based on their statistical typicality, which is computed as the ratio between phonological friends and orthographical neighbors. For instance, two-character kanji words with the character 神 in the first position have 84 orthographical neighbors, such as 神経 /siN-Kei/ (nerve), 神様 /kami-sama/(the god), and 神社 /ziN-jia/ (shrine). Among these orthographical neighbors, the word 神経 /siN-Kei/ has 71 phonological friends wherein 神 is similarly pronounced as /siN/. Therefore, it is categorized as an inconsistent but typical word as its phonological friends take the majority of the orthographical neighbors. By contrast, the words with 神 pronounced as /kami/, /ziN/, or /kaN/ in the same position take 13 of 84 positions, and are, hence, categorized as inconsistent-atypical words. Using this method, reading consistency can be captured in a numeric continuum (Figure 1a), wherein the value 1 represents a consistent word and a value less than 1 indicates an inconsistent word. Values come infinitely close to 0 with decreasing reading consistency. Thus, the degree of reading consistency is the statistical property of the O-P correspondences in Japanese kanji words. This statistical property is interpreted as the strength in weight of distributions, according to connectionist neurocomputational models (Coltheart et al., 2010; Plaut et al., 1996; Seidenberg & McClelland, 1989). Accordingly, reading consistency can be described as one aspect of statistical influence.

Consistency, frequency effect, and statistical learning

Fushimi et al. (1999) computed reading consistency and, thereby, demonstrated the *consistency effect*³ in oral reading tasks; accordingly, Japanese kanji words with high reading consistencies tend to generate short reaction times (RTs) and high reading accuracies. A *graded consistency effect* caused by a quasi-regular reading consistency structure—consistent word < inconsistent-typical word < inconsistent-atypical word—was observed in RT, as well. Additionally, the *frequency effect* was demonstrated within the same experimental word stimuli, that is, participants tended to perform more advantageously in high-frequency (HF) words in terms of RT and reading accuracy than in low-frequency (LF) words.

Consistency and frequency effects can be considered the consequences of statistical learning. The statistical learning paradigm was broadly discussed in the first-language (L1) acquisition of children and infants, such as early-age language input (Feldman et al., 2022; Richtsmeier & Goffman, 2017), syntax (Fisher et al., 2006), and phonotactic sequences (Richtsmeier et al., 2011; Richtsmeier & Goffman, 2023). The emergence of the HF advantage is considered a result of individuals' naturalistic exposure to the statistical characteristics of languages during early-age L1 acquisition and can be simulated in statistical learning experiments. Similarly, the quasi-regular structure in the reading consistency of Japanese kanji words is computed based on the lexical frequency in the corpora, which is a statistical property of the kanji word.

Consistency and frequency effects are not limited to Japanese; rather, they were first demonstrated in alphabetic writing systems. Research on English words reports a graded structure continuum comprising regular-consistent words, regular-inconsistent words, and exceptions from the highest to lowest reading accuracy. The

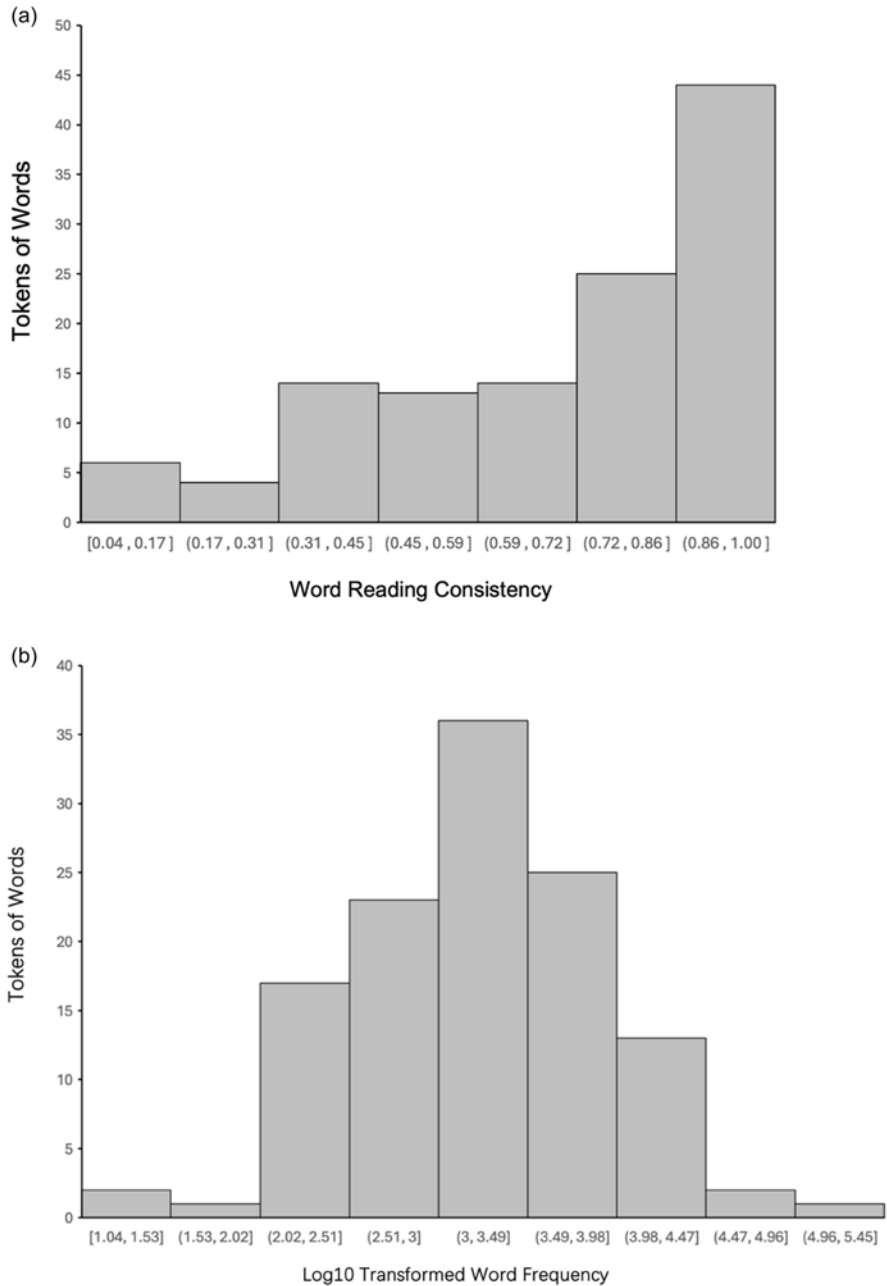


Figure 1. Distribution of word reading consistency (Figure 1a) and word frequency (Figure 1b) of the experimental materials. Figure 1a (top panel) shows the number of words distributed in each reading consistency interval. The reading consistency of word materials in Fushimi et al. (1999) was recomputed using the *Jōyō Kanji* and *Jukugo* database (Tamaoka et al., 2017). Figure 1b (bottom panel) shows the number of words distributed in each frequency interval. The lexical frequency was calculated using frequency lists in the *Chonagon* database, BCCWJ.

regularity-consistency effect has been verified in oral reading and lexical decision tasks (Andrews, 1982; Andrews et al., 2005; Glushko, 1979; Jared, 1997; Jared et al., 1990), as well. Logographic writing systems, for example, Chinese, report a similar regularity-consistency effect on the character–pronunciation correspondence in Mandarin (Chao et al., 2021; Chen et al., 2016; Lee et al., 2005; Li et al., 2020). The cross-linguistic generalization of the consistency effect provides further evidence of the consequence of statistical learning underlying word phonological processing.

Contribution of semantic memory

As mentioned earlier, whole-word-level reading is required to name an inconsistent Japanese kanji word. To perform whole-word-level reading, individuals are supposed to employ the lexical semantic knowledge stored in their semantic memory (Coltheart et al., 1993; Crisp & Lambon Ralph, 2006). The contribution of the semantic memory to word phonological processing is described as interactive access among semantic, orthographic, and phonological representational units in a *triangle neural connectionist network model* (Halai et al., 2018; Patterson & Lambon Ralph, 1999; Seidenberg & McClelland, 1989; Ueno et al., 2014). This has also been applied to the simulation of oral reading tasks by connectionist approach (Ijuin et al., 1996). Such models further explain how a Japanese kanji character is perceived at the whole-word level and the *graded consistency* effect is generated. Accordingly, stronger contributions of semantic memory are required to support whole-word-level reading for decreasing levels of the reading consistency of the target word.

However, when individuals encounter words with uncommon pronunciations, they make reading errors due to the default of the most typically used pronunciation. In such cases, a representative type of error—legitimate alternative reading of components (LARC) error—is generated. A LARC error refers to the inappropriate pronunciation of a target word but still correct once the constituent character is combined with other kanji as a different word (Fushimi et al., 1999; Fushimi et al., 2003; Fushimi et al., 2009). For instance, the mispronunciation of 神主 /kaN-nushi/ (lord god) as /siN-nushi/ is a LARC error since /siN/ is the correct pronunciation in words such as 神経 /siN-Kei/. Most of such errors are called *regularization errors* as the component characters' pronunciation tends to be wrongly regularized to the one with statistically higher reading consistency or lexical frequency. The presence of LARC errors further clarifies the influence of word statistical properties in phonological processing. For healthy Japanese L1 speakers, LARC errors are more likely to occur in low-frequency, than high-frequency, words because of limited statistical learning exposure. Nevertheless, L1 speakers rarely make reading errors, as semantic knowledge supposedly supports whole-word-level reading and compensates for the disadvantages of weak statistical properties (Ueno et al., 2014). For instance, although the word 金銭 /kiN-seN/ (money) is an inconsistent-atypical word, it has high lexical frequency. Therefore, the access to its semantic representation can be so automatic that reading errors are less likely to occur. In comparison, when naming an inconsistent-atypical word in the LF band, such as 毛皮 /ke-gawa/ (animal fur), a large magnitude of the semantic memory contribution is required to elicit the correct phonological representation. When the semantic contribution is unable to compensate for the least advantage of a statistical

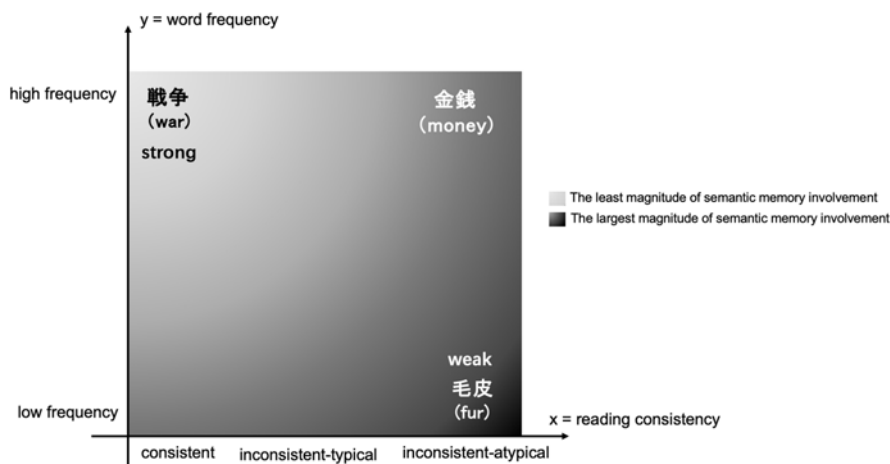


Figure 2. Illustration of the relationship between semantic memory and the emergence of consistency and frequency effects. The x- and y-axes represent the gradient reading consistency and word frequency variation, respectively. The words “strong” and “weak” refer to the engagement of statistical property (frequency and consistency). The intensity variation between gray and black colors within the square formed by x- and y-axes refers to the involvement of semantic memory, whose magnitude ranges from the least to the most. Therefore, when a word has high frequency and consistency, its statistical property is supposed to be the strongest so that the least magnitude of semantic memory is involved accordingly and vice versa.

property, LARC errors tend to appear. Hence, the emergence of frequency and consistency effects can be illustrated by the interaction between semantic memory and strength of two statistical properties (Figure 2). Further, this clarifies why healthy Japanese L1 speakers (Fushimi et al., 1999) are not highly sensitive to HF words’ quasi-regular structure in terms of reading accuracy and LARC errors and why their graded consistency effect is constrained to the LF band in RT. By contrast, Japanese semantic dementia patients having the same experimental material and design perform with great sensitivity to the quasi-regular structure, and their reading accuracy follows the consistent > typical > atypical pattern (Fushimi et al., 2009). Because of the neurological impairment in semantic memory processing, semantic dementia patients are assumed to name the Japanese kanji word with less reliance on semantic knowledge but heavy dependence on its statistical properties. Therefore, semantic dementia patients find it difficult to achieve the same reading accuracy quality as healthy Japanese natives in situations requiring strong semantic memory contributions. When the progressive deterioration in semantic memory becomes severe, their reading accuracy deteriorates due to the failure of word semantic activation (Fushimi et al., 2003; Fushimi et al., 2009).

Japanese Kanji learning for L1-Chinese speakers

Discussions on the phonological processing of Japanese kanji words in the context of L2 learning should consider factors related to L2 acquisition, such as L2 proficiency and L1 transfer. Earlier studies on Chinese L1 learners have extensively examined phonological, orthographical, and semantic parallels to Mandarin and

reported transfer effects in L2 Japanese word processing. For example, orthographic and semantic parallels, such as the character 表, convey the similarity of meanings, that is, “the cover” in both Chinese and Japanese. Effects in homophone-synonyms or cognates, such as the word 表象 (representation), occur due to orthographical/phonological similarities (Chiu, 2002a, 2002b). However, not all studies report the facilitating effect in L1 transfer (Fei, 2015; Kayamoto, 2000; Matsumi et al., 2014). In oral reading tasks, unlike readers with less proficiency, L2 learners with high proficiency were less affected by phonological/orthographical similarities in reading latency (Kayamoto, 2000, 2002). In lexical decision tasks involving manipulations of the switch in languages (Chinese–Japanese) and phonological similarities, intermediate learners were more likely to benefit from L1 phonological similarities, whereas advanced learners were more affected by L2 phonological knowledge (Fei, 2015). By contrast, negative transfer effects from L2 Japanese phonological similarities were observed in the L1 word processing of high-proficiency learners. In this case, the acquisition of Japanese phonological representations generated inhibitory effects on L1-Chinese phonological processing (Matsumi et al., 2014). These results imply the unstable nature of the facilitating effect of L1 transfer; it varies depending on the Japanese L2 proficiency levels of learners and on tasks with different requirements. Although the inhibitory/facilitative effects of orthographical/phonological similarities are assumed to influence the processing of Japanese kanji words, the processing can be modulated by the consequence of statistical learning in different L2 proficiency levels. Hence, the difficulties of the phonological processing of L2 Japanese kanji words cannot be comprehensively interpreted using the dichotomy of positive/negative transfer from L1.

The current study focuses on the impact of two statistical properties—that is, reading consistency and lexical frequency, rather than L1 transfers—on the oral reading performance of Japanese kanji words. It clarifies the intrinsic statistical properties of kanji words and provides insights into the potential difficulties in the phonological processing of L2 words.

Research questions and hypothesis

This study aims to illustrate the consistency and frequency effects in Chinese L1 learners of Japanese with varying L2 proficiencies. The following research questions (RQs) and hypotheses are tested:

Can the effects of reading consistency and lexical frequency in Japanese kanji words be demonstrated among L2-Japanese learners (RQ 1)? This study examined the two effects on RTs, reading accuracy, and LARC errors in L1-Chinese learners of Japanese. To ensure that participants had the requisite language proficiency to undertake an oral reading task, L2 learners with advanced proficiency were chosen in Experiment 1. If statistical learning is the foundation of language acquisition, then shorter RTs, higher reading accuracies, and fewer LARC errors would appear for words with high reading consistency and lexical frequency.

Does L2 learning proficiency influence consistency and frequency effects (RQ 2)? Experiment 2 was conducted to assess the two effects in lower-proficiency (intermediate) learners. To further refine our understanding of the relationship between L2 proficiency and word statistical properties, Japanese L2 proficiency

(advanced/intermediate) was introduced as a predictor of reading accuracy, using data from Experiments 1 and 2. Two different predictions were formulated.

(1) Higher proficiency learners are expected to exhibit a more pronounced graded consistency effect and its interaction with frequency compared with the lower-proficiency group. This prediction assumes that those with higher proficiency, due to longer exposure and more in-depth learning, will show better assimilation of the quasi-regular structure in reading consistency.

(2) On the contrary, high-proficiency learners might demonstrate reduced sensitivity to word statistical properties compared with low-proficiency learners who may exhibit increased sensitivity to reading consistency and frequency, reflecting their ongoing development in phonological and semantic knowledge. This prediction assumes that high-proficiency learners acquire more semantic knowledge of Japanese kanji than low-proficiency learners. To further explore this assumption, we compare the reading performances of intermediate learners with that of semantic dementia patients in Fushimi et al.'s (2009) study. This comparison enables us to examine whether the poor reading performance of lower-proficiency learners is caused at least partly by the learners' poor semantic knowledge of Japanese words. As intermediate learners are at a disadvantage in terms of their L2 proficiency, or their insufficient knowledge of the Japanese vocabulary, their reading accuracy was expected to be similar to the patterns reported by Fushimi et al. (2009).

Experiment 1

Experiment 1 investigated RQ 1. L2 learners with high Japanese proficiency were recruited as participants in this experiment.

Method

Participants

Postgraduate and undergraduate students majoring in Japanese or related subjects from a university in mainland China were recruited through the social network platform WeChat. Although all participants were studying at the same university, they came from different districts in mainland China where Standard Mandarin is spoken natively.⁴ All participants were required to have passed N1, the highest level of the Japanese Language Proficiency Test (JLPT), within the past 5 years and to provide their scores in a survey.⁵ The survey was administered to confirm that all of them were Japanese as a foreign language learners and started formal Japanese learning while being enrolled as undergraduates. In Experiment 1, 33 university students ($M_{\text{age}} = 22.71$ years, standard deviation (SD) = 1.67; 6 men and 27 women) were recruited. Two of the participants were fourth-year undergraduate students majoring in Japanese language and literature, whereas the remainder were postgraduates majoring in Japanese language, education, economics, or social science. Their average score on JLPT-N1 was 149.86 points (SD = 20.34, max = 180, min = 113). The participants' mean duration of Japanese learning was 63.43 months (SD = 14.81, max = 96, min = 40). They had little experience studying Japanese overseas during their university education, except for nine postgraduate students who had spent a maximum of 12 months in Japan. Five participants attended the pilot study, and the

remaining 28 attended the formal experiment. The data from the formal experiment alone were included in the analysis.

Materials and design

One hundred twenty word stimuli divided into three consistency groups, consistent, inconsistent-typical, and inconsistent-atypical, were obtained from Fushimi et al. (1999). Forty kanji words in each group were divided into HF and LF bands. Hence, 20 words were allocated to each of the six frequency-consistency stimulus blocks.

Procedure

The experiment was conducted in a university classroom in mainland China, with participants attending in person and being tested individually. Each was equipped with Japanese-made Century-SSS108VR earphones, featuring a built-in microphone for voice recording. The session ran on Psychopy 3.0 software (Peirce et al., 2019). At the start, participants pressed the space bar in response to a fixation point on the screen to initiate each trial. A set of 120-word stimuli was randomly presented, each appearing for 5,000 ms followed by automatic progression to the next word. Participants were instructed to read each word aloud as quickly and accurately as possible within the 5,000 ms. A practice phase of 10 trials preceded the experimental phase, with no answers displayed for any trial. The entire experiment lasted approximately 15 minutes.

Analysis

The word statistical data used in Fushimi et al. (1999) were based on a corpus published in 1970 by the National Language Research Institute. The present study updated psycholinguistic properties of word stimuli according to the analysis of current corpora (Table 1).

Lexical frequency

Word stimuli were reassessed using the frequency lists published in *Chunagon*, the Balanced Corpus of Contemporary Written Japanese (BCCWJ), which was developed by the National Institute for Japanese Language and Linguistics in 2021.⁶ Since some words in the HF group have become less commonly used, minor adjustments were made according to the updated frequency rankings, resulting in three to four words being shifted between the original HF and LF bands. Figure 1b illustrates the distribution of word frequencies. After the adaptation, significant HF–LF word band differences were confirmed in consistent ($t(31.85) = 8.45, p < 0.01, d = 2.67$), inconsistent-typical ($t(37.83) = 8.32, p < 0.01, d = 2.63$), and inconsistent-atypical ($t(36.19) = 8.28, p < 0.01, d = 2.62$) word groups.

Word reading consistency

Reading consistency was recomputed using the *Jōyō Kanji* and *Jukugo* database (Tamaoka et al., 2017). This database was selected because (1) it provides a list of

Table 1. Characteristics of word stimuli in Experiment 1

Characteristic	Consistent		Inconsistent- typical		Inconsistent- atypical	
	HF	LF	HF	LF	HF	LF
No. of on-reading words	19	20	20	19	11	3
No. of L1 orthographic similar words	19	13	18	10	11	6
Mean word frequency ^a	3.77	2.66	3.72	2.87	3.74	2.45
Mean no. of orthographical neighbors	Left	27.35	26.40	40.05	25.55	43.10
	Right	27.25	26.70	38.75	33.70	49.05
Mean no. of phonological friends	Left	27.20	26.40	29.50	18.30	14.95
	Right	27.25	26.70	30.65	23.50	24.10
Mean word consistency	1.00	1.00	0.77	0.72	0.45	0.39

Note: ^aWord frequency is the transformed value of Log10 from the frequency list of *Chunagon*, BCCWJ.

commonly used Japanese kanji characters and modern kanji idioms based on the corpus derived from the Mainichi newspaper articles published between 2000 and 2010. (2) The orthographical and phonological neighbors of two-character kanji words (*Jukugo*) can be accessed directly through an interactive website (www.kanjidatabase.com). Based on the corpus, the numbers of phonological friends and orthographic neighbors of the right- and left-side characters for each word stimulus were calculated. Word reading consistency was computed as the number of phonological friends/the number of orthographic neighbors (Fushimi et al., 1999; Fushimi et al., 2009), and the average value of the reading consistencies for the left- and right-side characters was counted as word reading consistency. A one-way analysis of variance (ANOVA) was conducted to confirm the differences among the three grades of reading consistency proposed by Fushimi et al. (1999). The analysis revealed significant differences among the three consistency groups ($F(2, 117) = 217.80$, mean-squared error (MSE) = 0.02, $p < 0.01$, $\eta^2_p = 0.79$), and significant differences between inconsistent-atypical and inconsistent-typical according to post hoc comparisons (mean difference = -0.33, standard error (SE) = 0.03, $t(2, 57) = -9.69$, $p < 0.01$). These results indicate that the classification of reading consistency provided by Fushimi et al. (1999) is still applicable to the current corpus.

Tests of consistency and frequency effects

Mixed-effects modeling was used to examine the main effects of consistency and frequency, as well as their interactions. Given that interaction effects provide a more comprehensive understanding of participants' sensitivity to the word statistical prosperities, models that included the significant interaction terms were prioritized in the model selections. Baseline models were first constructed to assess potential confounding effects from two L1-transfer factors⁷: L1-orthographical similarity

(whether the combination of the two Japanese kanji characters exists as a word in Chinese) and pronunciation type (whether it is an on-reading word or not). If the main effect of the two factors were non-significant, they were included as interactions with consistency and frequency in the subsequent model constructions. Consistency and frequency were also included as random slopes in the random effects to account for individual variability. Within this modeling framework, a traditional stepwise regression approach utilizing backward elimination was employed. Models that successfully converged and avoided boundary fit issues were then compared using a likelihood ratio test via the *anova* function in R (Kuznetsova et al., 2020) for the optimal. Results of Type III Wald chi-square (Schad, 2020) for the optimal models were reported in the main text, using *car* package in R (Fox & Weisberg, 2023). Detailed results of mixed-effect model analyses including coefficients and random effects structures⁸ are supplemented in Appendix A.

To further examine the *graded consistency effect* and its interaction with lexical frequency, pairwise comparisons were conducted using the *emmeans* package in R (Lenth, 2023), with detailed results presented in Appendix B. To ensure the robustness of the best-fitting model, secondary analyses were conducted using continuous-variable models, representing frequency and consistency on a numerical scale using data in Table 1. Rather than shifting between discrete category levels, frequency, and consistency were treated as continuous variables, allowing the impact of gradual, incremental changes in statistical properties to be captured. The continuous variables were directly applied to the best-fitting model to streamline the analysis while preserving the validated model structure.

Results

Reaction time

The RT for each word stimulus was measured from the visual presentation to the oral response onset, using Chronset (Roux et al., 2017), an automated speech detection tool. Mean RTs for each word class are summarized in Table 2. Approximately 14.8% of responses, including incorrect responses and no-responses, were invalidated and excluded from further analysis. The remaining valid responses were analyzed using a linear mixed-effects regression model implemented in the *lmerTest* package in R (Kuznetsova et al., 2017). The effects of L1-orthographical similarity ($\chi^2 = 0.38$, $df = 1$, $p = 0.53$) and pronunciation types ($\chi^2 = 2.58$, $df = 1$, $p = 0.11$) were not significant, as shown in Table 3. Model comparisons (Kuznetsova et al., 2020) indicated that model $RT \sim \text{Frequency} \times \text{Consistency} + (1 + \text{Frequency} \mid \text{participant}) + (1 \mid \text{word})$ provided the lowest Akaike information criterion scores and significantly accounted for variations ($\chi^2 = 17.17$, $df = 2$, $p < 0.01$), making it the optimal one for RT (Appendix A, Table A1). Type III chi-square test (Fox & Weisberg, 2023) showed a significant main effect of lexical frequency ($\chi^2 = 53.34$, $df = 1$, $p < 0.01$), reading consistency ($\chi^2 = 23.11$, $df = 2$, $p < 0.01$), and significant interactions ($\chi^2 = 13.06$, $df = 2$, $p < 0.01$). Pairwise comparisons (Lenth, 2023) further clarified that HF words consistently produced shorter RTs compared with LF words across all consistency levels. And a graded

Table 2. Mean values of RT, error rates, and LARC error responses for each word class in Experiments 1 and 2

Consistency	High Frequency			Low Frequency		
	Consistent	Typical	Atypical	Consistent	Typical	Atypical
Experiment 1						
RT (ms)	1,019.05	1,131.13	1,117.08	1,250.54	1,383.89	1,598.75
(SD)	(223.39)	(223.75)	(194.79)	(252.15)	(300.26)	(344.39)
Error rate (%)	1.25	3.75	8.04	4.29	21.43	45.86
(SD)	(3.23)	(4.84)	(6.43)	(4.85)	(11.04)	(16.28)
LARC error (%)	0.00	2.69	4.82	0.54	11.79	31.01
(SD)	(0.00)	(4.61)	(5.18)	(1.57)	(7.48)	(15.45)
Experiment 2						
Error Rate (%)	7.44	15.66	21.75	14.69	35.88	70.80
(SD)	(10.56)	(12.09)	(13.81)	(13.53)	(18.12)	(18.22)
LARC Error (%)	0.97	6.61	10.70	3.72	15.18	37.95
(SD)	(2.01)	(6.11)	(8.06)	(4.08)	(8.50)	(14.67)

Note: LARC = legitimate alternative reading of components; RT = reaction time; SD = standard deviation; Typical = inconsistent-typical; Atypical = inconsistent-atypical.

Table 3. Type III Wald chi-square test results for Experiments 1 and 2

Outcome Variable	Effect	Chisq	df	p	
Reaction Time (Experiment 1)	Intercept	721.09	1	< 0.001	***
	Frequency	53.34	1	< 0.001	***
	Consistency	23.11	2	< 0.001	***
	Interaction	13.06	2	< 0.01	**
Reading Accuracy (Experiment 1)	Intercept	164.89	1	< 0.001	***
	Frequency	53.36	1	< 0.001	***
	Consistency	46.83	2	< 0.001	***
Reading Accuracy (Experiment 2)	Intercept	50.99	1	< 0.001	***
	Frequency	46.70	1	< 0.001	***
	Consistency	51.02	2	< 0.001	***
	Interaction	13.42	2	< 0.01	**

Note: Formula for reaction time: $RT \sim \text{Frequency} \times \text{Consistency} + (1 + \text{Frequency} | \text{subject}) + (1 | \text{word})$, the number of observations is 2,832; Formula for reading accuracy in Experiment 1: $\text{Correct_Incorrect} \sim \text{Frequency} + \text{Consistency} + (1 | \text{subject}) + (1 | \text{word})$, the number of observations is 3,332; Formula for reading accuracy in Experiment 2: $\text{Correct_Incorrect} \sim \text{Frequency} \times \text{Consistency} + (1 + \text{Consistency} | \text{subject}) + (1 | \text{word})$, the number of observations is 3,706; ** $p < 0.01$; *** $p < 0.001$.

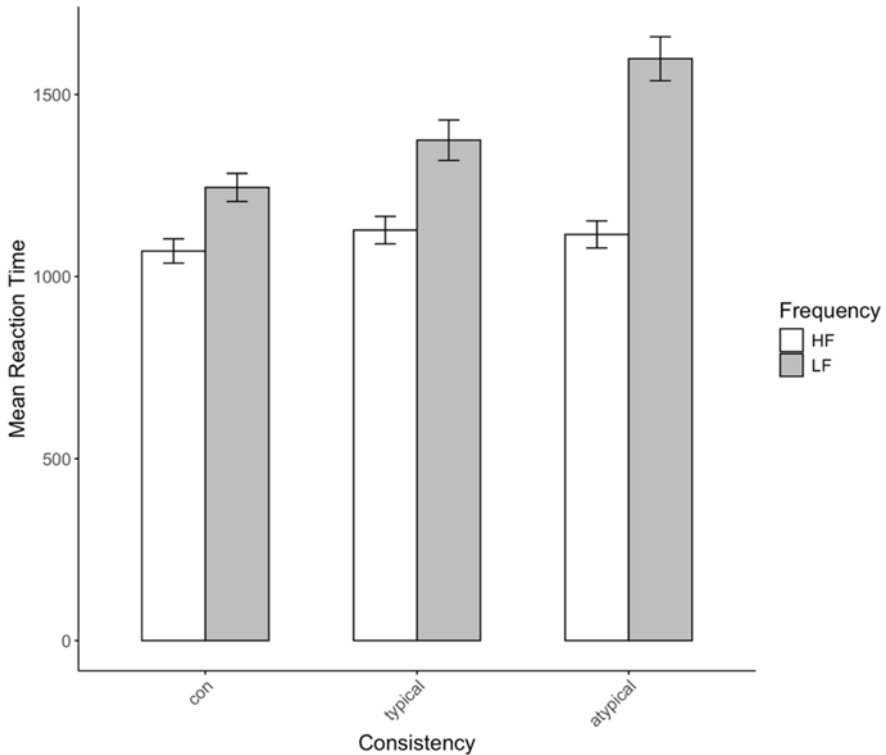


Figure 3. Reaction time with 95% confidence intervals in experiment 1. con = consistent word; typical = inconsistent-typical word; atypical = inconsistent-atypical word. HF and LF represent high- and low-frequency words, respectively. Error bars indicate 95% confidence intervals for within-subject comparisons (Cumming & Finch, 2005).

consistency effect in RT (consistent < typical < atypical) was observed within the LF band (Appendix B, Table B1), as shown in Figure 3.

Reading accuracy

Table 2 and Figure 4 present the mean reading error rates (ERs) for each word class. Reading accuracy was coded as correct or incorrect and analyzed using a generalized linear mixed model via the *lme4* package in R (Bates et al., 2014). Among all responses, 0.8% were invalidated due to irregularities during the program's execution.⁹ The procedure for model selections mirrored that used for RT. The baseline model showed no significant main effects of pronunciation types ($\chi^2 = 0.24$, $df = 1$, $p = 0.62$) or L1-orthographical similarities ($\chi^2 = 0.21$, $df = 1$, $p = 0.65$). Model comparisons (Kuznetsova et al., 2020) showed that *Correct/Incorrect ~ Frequency + Consistency + (1 | participant) + (1 | word)* provided the best fit for reading accuracy, as adding interactions did not significantly improve model fitness (Appendix A, Table A4). Results revealed a significant main effect of lexical frequency ($\chi^2 = 53.36$, $df = 1$, $p < 0.01$) and reading consistency

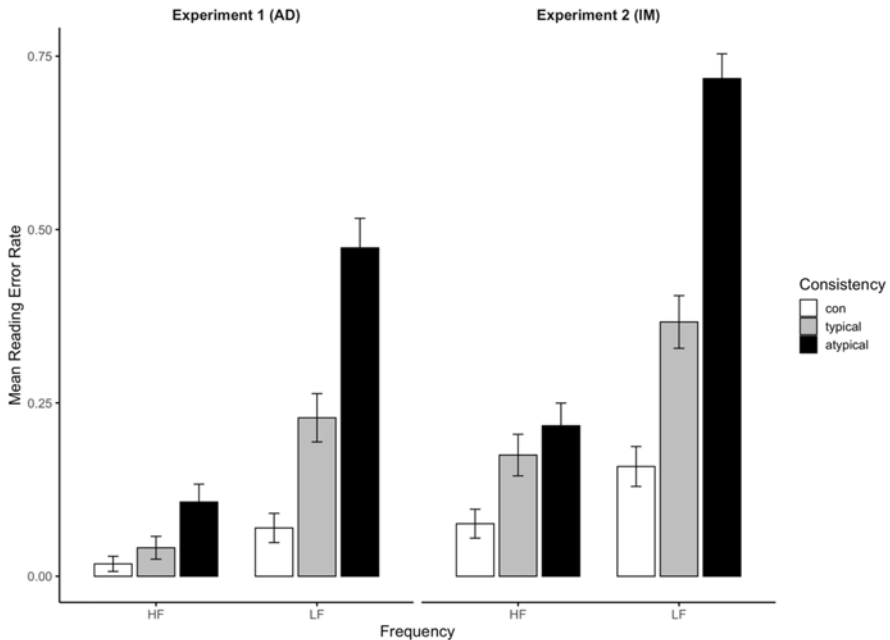


Figure 4. Mean reading error rates of Experiments 1 and 2. AD and IM represent advanced and intermediate learner group, respectively. con = consistent word, typical = inconsistent-typical word, atypical = inconsistent-atypical word; HF = high-frequency words, LF = low-frequency words. Error bars are 95% confidence intervals for within-subject comparisons (Cumming & Finch, 2005).

($\chi^2 = 46.83$, $df = 2$, $p < 0.01$). The pairwise comparisons (Lenth, 2023) further confirmed the *graded consistency effect* (consistent < typical < atypical), indicating a higher likelihood of incorrect responses in words with lower reading consistency (Appendix B, Table B2).

To further evaluate the robustness of these findings in the RT and reading accuracy, continuous-variable models were conducted with the same structure of best-fit models. Both consistency and frequency were treated as numeric variables and subjected to a z-score standardization. The results were consistent with the optimal model on RT and reading accuracy (Appendix A, Tables A3 and A6).

LARC errors

LARC errors are inappropriate responses to target two-character words. However, the response of a single character can be a correct pronunciation when it is combined with other kanji characters (Fushimi et al., 1999; Fushimi et al., 2003; Fushimi et al., 2009; Patterson et al., 1995). In this study, LARC errors were defined without regard to the kanji character's position within the word (Fushimi et al., 2009; Patterson et al., 1995).¹⁰

Participants' responses were classified as correct, LARC error, or other types of errors, and were examined using a multinomial logistic regression model

Table 4. Results of multinomial logistic regression analysis on LARC error in Experiments 1 and 2

Experiment 1				
Fixed effects	β	SE	z	$p (> t)$
Intercept	-2.85	0.19	-14.91	< 0.001 ***
Frequency	2.32	0.21	10.83	< 0.001 ***
Consistency 1 (con vs. atyp)	-14.92	0.36	-41.71	< 0.001 ***
Consistency 2 (typ vs. atyp)	-0.73	0.32	-2.25	< 0.05 *
Frequency $\times$ Consistency 1	9.88	0.36	27.62	< 0.001 ***
Frequency $\times$ Consistency 2	-0.67	0.36	-1.84	0.07
Experiment 2				
Fixed effects	β	SE	z	$p (> t)$
Intercept	-2.05	0.13	15.22	< 0.001 ***
Frequency	2.33	0.17	13.86	< 0.001 ***
Consistency 1 (con vs. atyp)	-3.20	0.59	-5.38	< 0.001 ***
Consistency 2 (typ vs. atyp)	-0.49	0.21	-2.32	< 0.05 *
Frequency $\times$ Consistency 1	-0.29	0.64	-0.45	0.65
Frequency $\times$ Consistency 2	-1.18	0.26	-4.54	< 0.001 ***

Note: Formula: LARC response $\sim$ Frequency $\times$ Consistency; con = consistent, typ = inconsistent-typical, atyp = inconsistent-atypical; SE = standard error, SD = standard deviation; * $p < 0.05$; *** $p < 0.001$.

implemented with the *nnet* package in R (Ripley et al., 2016). As shown in Table 4, the model that included the consistency $\times$ frequency term showed that LF words, than HF words, were significantly more likely to result in LARC errors ($\beta = 2.32$, $SE = 0.21$, $z = 10.83$, $p < 0.01$). Significant differences were also found between consistent and inconsistent-atypical words ($\beta = -14.92$, $SE = 0.36$, $z = -41.71$, $p < 0.01$), and between inconsistent-typical and inconsistent-atypical words ($\beta = -0.73$, $SE = 0.32$, $z = -2.25$, $p < 0.05$). The interaction effect further revealed that word frequency significantly modulated the difference between consistent and inconsistent-atypical words ($\beta = 9.88$, $SE = 0.36$, $z = 27.62$, $p < 0.01$) but not the differences between inconsistent-typical and inconsistent-atypical words ($\beta = -0.67$, $SE = 0.36$, $z = -1.84$, $p = 0.07$). Pairwise comparisons (Lenth, 2023) further indicated a significant graded consistency effect in the LF band. However, the frequency effect was only significant in the inconsistent-typical and inconsistent-atypical word groups (Appendix B, Table B3).

Experiment 2

Experiment 2 was aimed to address RQ 2 by engaging intermediate-level L2 learners. The oral reading task and the data analysis followed the same procedure utilized in Experiment 1.

Method

Participants

Experiment 2 was conducted as an online experiment.¹¹ Participants, all undergraduates majoring in Japanese from various universities in mainland China, were recruited using the same procedure as in Experiment 1. Although participants were informed that having a certain degree of Japanese proficiency was a prerequisite to completing the experimental task, a primary criterion for inclusion was their status as undergraduates majoring in Japanese. Their Japanese proficiency as an intermediate level was further verified through an identical survey, as used in Experiment 1. Ultimately, 38 university students ($M_{\text{age}} = 20.53$, $SD = 1.09$; 6 men, 32 women) participated. These students, ranging from second to fourth-year students, were primarily majoring in Japanese language and literature and had minimal Japanese learning experience prior to university enrollment. Fifteen participants had taken the JLPT-N1, achieving an average score of 115.93 ($SD = 20.24$, $\text{max} = 140$, $\text{min} = 79$), with four scoring below 100 points. Eight participants had attempted the JLPT-N2, with an average score of 124.87 ($SD = 20.03$, $\text{max} = 148$, $\text{min} = 92$), and three of these did not pass, scoring below 100. The average duration of studying Japanese among the participants was 31.93 months ($SD = 8.06$, $\text{mix} = 12$, $\text{max} = 59$). Most participants had little to no experience studying Japanese abroad; however, five had lived in Japan for less than six months, and one had lived there for nearly twelve months.

Four participants attended a pilot study, while the remaining 34 were included in the formal experiment for subsequent analysis. The materials and design used in Experiment 2 were identical to those of Experiment 1.

Procedure

The oral word reading task was conducted online via ZOOM. Before the experiment began, participants were required to prepare a PC equipped with a built-in microphone. They were also instructed to ensure their environment was quiet, and their devices were set up for Internet access and ZOOM registration. This task utilized the same software and program as Experiment 1. Once the experiment commenced, the experimenter presented the word stimuli using ZOOM's screen-sharing function, and participants' responses were recorded using the platform's recording function throughout the session.

Analysis

The data analysis procedure for Experiment 2 was similar to that of Experiment 1. However, due to network transmission delays, RTs could not be captured. Additionally, the data from two participants were not fully recorded because of unexpected network server errors.

As some participants had not passed JLPT, this study examined L2 proficiency levels by assessing the reading ERs of word stimuli with intermediate levels of difficulty. The level of L2 learning difficulties of the 120-word stimuli was assessed using the Japanese Language Reading Tutorial System, which was developed by the University of Tsukuba.¹² The system indicated that 20 words were within the level

N1, 43 words within N2–N3, and 15 words within N4–N5, and 42 words were not registered within the JLPT levels. Since words in the N2–N3 level are generally considered to be in the intermediate level, 24 words were selected from the N2–N5 difficulty level as testing stimuli. Reading ER of a 12-word set with N2–N3 levels of difficulty was first examined, comprising six HF and six LF words. The average ER of all the participants was 0.29 (SD = 0.18, min = 0.00, max = 0.75). Subsequently, the remaining set of words, comprising six HF words within the N4–N5 difficulty level and six LF words within the N2–N3 difficulty level, were tested, and the average ER was 0.28 (SD = 0.18, max = 0.67, min = 0.00). ERs exceeding $M+2SD$ in both sets of words (over 0.65) were considered outliers. As a result, one participant who had generated ERs of 0.75 and 0.67 in the first and second sets, respectively, was excluded. Consequently, the reading accuracy of 31 participants was analyzed in the mixed-effects model.

Results

Reading accuracy

Table 2 and Figure 4 display the average percentage of reading ERs in each word class. Due to irregularities during program execution, 0.4% of all responses were deemed invalid. The analysis was conducted using a mixed-effects model similar to Experiment 1. Test of the baseline model showed that the main effect of pronunciation types ($\chi^2 = 0.04$, $df = 1$, $p = 0.85$) and L1-orthographical similarities ($\chi^2 = 1.26$, $df = 1$, $p = 0.26$) were insignificant. Model comparisons (Kuznetsova et al., 2020) of interaction models (details in Appendix A, Table A7) indicated that *Correct/Incorrect* $\sim$ *Frequency* $\times$ *Consistency* + (*1* + *Consistency* | *participant*) + (*1* | *word*) provided the best fit ($\chi^2 = 11.60$, $df = 1$, $p < 0.01$). Type III Wald Chi-square tested (Fox & Weisberg, 2023) revealed significant effects of lexical frequency ($\chi^2 = 46.70$, $df = 1$, $p < 0.01$), reading consistency ($\chi^2 = 51.02$, $df = 2$, $p < 0.01$), as well as their interaction ($\chi^2 = 13.42$, $df = 2$, $p < 0.01$). The continuous-variable model (Appendix A, Table A9) corroborated these findings. Pairwise comparisons (Lenth, 2023) further confirmed a significant frequency effect within each consistency level and a graded consistency effect in the LF band (Appendix B, Table B4). The consistency effect was also significant across the HF word band, except for the difference between inconsistent-typical and inconsistent-atypical word groups.

LARC error

The multinomial regression analysis (Ripley et al., 2016) demonstrated a significantly higher likelihood of LARC error responses in the LF band compared to the HF band ($\beta = 2.33$, SE = 0.17, $z = 13.86$, $p < 0.01$). Significant differences were observed between consistent and inconsistent-atypical words ($\beta = -3.20$, SE = 0.59, $z = -5.38$, $p < 0.01$), as well as between inconsistent-typical and inconsistent-atypical words ($\beta = -0.49$, SE = 0.21, $z = -2.32$, $p < 0.05$). The interaction between word frequency and reading consistency was significant, with the differences between inconsistent-typical and inconsistent-atypical words significantly interacting with word frequency ($\beta = -1.18$, SE = 0.26,

Table 5. Type III Wald chi-square test for the impact of L2 proficiency, consistency, and frequency effect on reading accuracy (Experiments 1 and 2)

Effect	<i>Chisq</i>	<i>df</i>	<i>p</i>	
Intercept	157.28	1	< 0.001	***
Frequency	56.29	1	< 0.001	***
Consistency	40.99	2	< 0.001	***
Proficiency	25.80	1	< 0.001	***
Frequency × Consistency	3.43	2	0.18	
Frequency × Proficiency	3.03	1	0.08	
Consistency × Proficiency	0.05	2	0.97	
Frequency × Proficiency × Consistency	12.17	2	< 0.01	**

Note: Formulas: Reading Accuracy ~ Frequency × Consistency × L2 Proficiency + (1 + Consistency | subject) + (1 | word); the number of observations is 7,038. ** $p < 0.01$; *** $p < 0.001$.

$z = -4.54, p < 0.01$). Pairwise comparisons (Lenth, 2023) confirmed a significant frequency effect across all consistency groups, and a significant graded consistency effect was captured in the LF band (Appendix B, Table B5).

Impact of L2 learning proficiency on reading accuracy

To investigate the impact of L2 proficiency, the reading accuracy data from advanced learners in Experiment 1 was compared with that of intermediate learners in Experiment 2. Word frequency, reading consistency, and L2 proficiency were submitted as fixed effect factors into the mixed-effect modeling. Model comparisons (Kuznetsova et al., 2020) revealed that a three-way interaction model, *Correct/Incorrect ~ Frequency × Consistency × L2 Proficiency + (1 + Consistency | subject) + (1 | word)*, provided the best-fit information criterion ($\chi^2 = 21.41, df = 5, p < 0.01$, see details in Appendix C, Table C1). As shown in Table 5, this model highlighted significant main effects for L2 proficiency ($\chi^2 = 25.80, df = 1, p < 0.01$), lexical frequency ($\chi^2 = 56.29, df = 1, p < 0.01$), and reading consistency ($\chi^2 = 40.99, df = 2, p < 0.01$). Significant three-way interactions between the three fixed factors were also observed ($\chi^2 = 12.17, df = 2, p < 0.01$). These findings were also supported by the continuous-variable model (Appendix C, Table C3), indicating a robust three-way interaction effect. Pairwise comparisons (Lenth, 2023) confirmed the significant impact of L2 proficiency across each frequency-consistency block, with a notable graded consistency effect in the LF band observed in both learner groups (Appendix C, Table C4). Additionally, the intermediate learner group exhibited greater mean differences among the three levels of reading consistency within the LF band, underscoring the nuanced impact of proficiency on reading performance.

Other types of errors in experiments 1 and 2

In addition to LARC errors, reading errors were categorized as one-character errors, visual-semantic errors, disfluency errors, and “others,” which included responses

that did not fit the previous categories. Table 6 presents the mean rate of each error type.¹³ As shown, one-character errors—where one kanji is pronounced correctly within a word while the other is incorrect but not classified as a LARC error—were notably high in the inconsistent-typical and inconsistent-atypical LF bands, both at 11.94%, in Experiment 2.

These errors highlight incomplete phonological knowledge at the sub-word level. Visual-semantic errors occur when responses share visual and/or semantic similarities with the target words, such as pronouncing 夜間 /ya-kaN/ as 夜中 /yo-naka/. Disfluency errors, involving inappropriate pauses or elongated vowels, reflect challenges in automatic phonological processing. The “others” category, comprising “do-not-know” responses and other phonological mistakes, was particularly prevalent in the inconsistent-typical and inconsistent-atypical LF bands, indicating a significant lack of phonological knowledge among intermediate learners.

Comparison with Japanese semantic dementia patients

To investigate whether the reading performance of low-proficiency learners can be partly due to their limited semantic knowledge of Japanese words, a three-way mixed-design ANOVA (group $\times$ frequency $\times$ consistency) was conducted. The analysis compared the reading accuracy of two groups—the intermediate learners in Experiment 2 and 14 observations from Japanese semantic dementia patients reported by Fushimi et al. (2009, Table 5, p. 1066), using the *lme4* and *afex* packages in R (Bates et al., 2014; Singmann et al., 2015). Statistical analysis revealed a significant effect of lexical frequency ($F(1, 43) = 241.44$, $MSE = 93.16$, $p < 0.01$), and reading consistency ($F(2, 86) = 166.15$, $MSE = 122.85$, $p < 0.01$). Although the main effect of group differences was insignificant ($F(1, 43) = 1.76$, $MSE = 865.02$, $p = 0.04$), the interaction between consistency and frequency effect proved to be significant ($F(2, 86) = 21.92$, $MSE = 70.19$, $p < 0.01$). Post-hoc comparison (Lenth, 2023) further indicated robust graded consistency effects in each frequency band for both groups (Appendix D, Table D2), as shown in Figure 5.

Discussion

The results of Experiments 1 and 2 collectively demonstrate consistency and frequency effects on reading performance among advanced and intermediate L2 learners of Japanese. As expected from the reasoning for RQ 1, the consistency and frequency effects can be extended to L2 learners. Generally, words with lower reading consistency and lexical frequency led to longer RTs (Experiment 1), lower reading accuracy, and higher rates of LARC errors. Continuous-variable model analyses confirmed these findings by capturing the impact of reading consistency and lexical frequency quantitatively.

For RQ 2, this study indeed demonstrated the influence of L2 learning proficiency, which leads to different patterns of the two effects in reading accuracy. The results tend to verify the second prediction: that low-proficient learners exhibit increased sensitivity to reading consistency and frequency, particularly in the quasi-regular structure reflected in the graded consistency effect. For advanced learners

Table 6. Mean rates of different types of errors for each word class in Experiments 1 and 2

	High frequency			Low frequency		
	Consistent	Typical	Atypical	Consistent	Typical	Atypical
Experiment 1						
Correct answers	98.75	96.25	92.14	95.54	78.57	54.32
LARC errors	0.00	2.68	4.82	0.10	11.79	31.39
One-character errors	0.89	0.54	1.25	2.50	4.29	3.38
Visual-semantic errors	0.36	0.36	1.43	1.43	2.32	3.04
Disfluency errors	0.00	0.00	0.00	0.00	0.00	0.00
Others	0.00	0.18	0.71	0.36	3.39	7.50
Experiment 2						
Correct answers	92.56	84.34	78.25	85.31	64.12	29.20
LARC errors	0.97	6.61	10.70	14.69	15.18	37.95
One-character errors	5.32	4.84	6.45	8.39	11.94	11.94
Visual-semantic errors	0.17	1.45	6.48	1.45	2.42	2.27
Disfluency errors	0.00	0.65	0.32	0.81	0.32	1.29
Others	1.29	3.73	1.66	3.40	5.71	11.43

Note: LARC = legitimate alternative reading of components; Typical = Inconsistent-typical; Atypical = inconsistent-atypical.

(Experiment 1), the interaction effect and the graded consistency effect, limited to the LF band, were notably detected in RT, whereas no interaction effect was significant in reading accuracy. This pattern shared similarities to native Japanese healthy adult speakers (Fushimi et al., 1999), indicating that advanced learners tended to be more subject to the statistical properties in terms of processing automaticity, rather than accuracy. In contrast, intermediate learners (Experiment 2) were more sensitive to the two statistical properties and their interactions. Results of the three-way interaction model, as well as pairwise comparisons, further confirmed a robust graded consistency effect in reading accuracy. This can be due to the lack of consolidated word phonological or semantic knowledge for L2 low-proficient learners. The analysis of other error types implied the former: intermediate learners produced more one-character errors and “other” errors (including no-response errors) compared with high-proficient learners. Moreover, a high degree of approximation in reading accuracy patterns between intermediate learners and Japanese SD patients (Fushimi et al., 2009) adds further support to the latter. Although the difference between intermediate learners and semantic dementia patients was not statistically significant, the graded consistency effect remained robust across each frequency band, suggesting that both groups struggle to activate robust semantic knowledge. Collectively, these findings indicate that less proficient learners rely heavily on statistical properties and occasionally lack sufficient phonological/semantic knowledge of Japanese kanji.

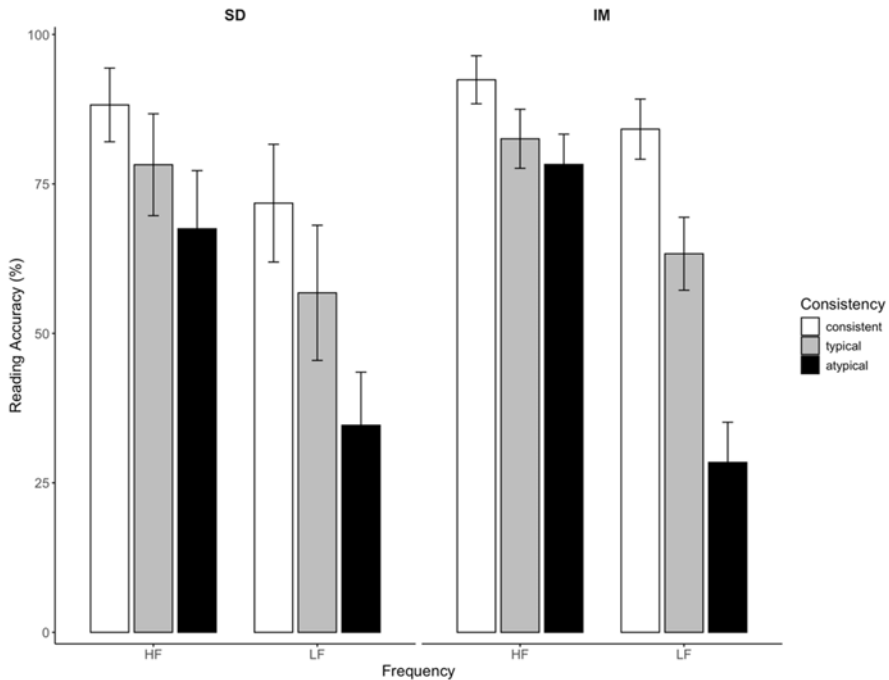


Figure 5. Mean reading accuracy of L2 intermediate learners and Japanese semantic dementia patients. The figure depicts the mean reading accuracy for each word class. SD = Japanese semantic dementia patients in the study by Fushimi et al. (2009); IM = second-language (L2) intermediate learners in Experiment 2; con = consistent word, typical = inconsistent-typical word, atypical = inconsistent-atypical word. HF and LF represent high- and low-frequency words, respectively. Error bars indicate 95% confidence intervals for within-subject comparisons (Cumming & Finch, 2005).

Finally, the frequency-modulated consistency effect, that is, the LF-constrained graded consistency effect, was observed both in reading accuracy and LARC errors across each L2 proficiency group. This is consistent with the theoretical models, as shown in Figure 2 and other connectionist neural network, where HF exposure strengthens phonological-semantic connections, minimizing the graded consistency effect due to the compensation of our semantic knowledge.

General discussion

L2 phonological processing and statistical learning

This study demonstrated that two statistical properties, that is, lexical frequency and reading consistency, in Japanese kanji words have a robust impact on word phonological processing for L2 Japanese learners. It further suggests that statistical learning is an essential capability that can provide insights into both L2 word phonological processing and L2 acquisition.

Regarding the former consideration, L2 word phonological processing, the two experiments in this study indicate that similar to native speakers of Japanese, L2 learners are disadvantaged in terms of processing accuracy and automaticity

(reflected in RTs) for words with weak statistical properties. This phenomenon can be explained as the interaction between word statistical property and semantic memory. This finding is of particular significance for participants who are L1-Chinese speakers and whose writing system shares similarities with the Japanese kanji system. Although earlier studies extensively discuss the facilitating/inhibitive effects of L1 transfer, they have minimal focus on the commonalities with L1 speakers in Japanese word processing. However, the results of the mixed-effects model analysis indicated that L1-orthographical similarities and on-reading words had only minimal impact, at least in this study, whereas lexical frequency and reading consistency had significant effects. As this study did not focus on L1 transfer and used the word stimuli that were initially designed for Japanese natives, relevant materials were probably not suitable for detecting L1 transfer effects.

Regarding the second consideration, L2 acquisition, this study implies that L2 learners may encounter difficulties in word phonological processing on words with inherently weak statistical properties, that is, when the constituent character of a kanji word has multiple pronunciations or low lexical frequency. Statistical learning is assumed not to be consciously captured in explicit learning contexts, as observed in L1 acquisition through natural linguistic input/output. This study provides evidence of the impact of such statistical learning on L2 acquisition.

L2 proficiency features in naming Japanese Kanji words

L2 learners with different proficiency levels have different sensitivities toward word statistical properties. This phenomenon indicates that statistical learning in L2 acquisition is a perpetually dynamic process. With the enhancement of L2 proficiency, the performance of word phonological processing converges to that of Japanese native speakers. This can be elaborated, as suggested by the triangle neural connectionist network model (Halai et al., 2018; Patterson & Lambon Ralph, 1999; Seidenberg & McClelland, 1989; Ueno et al., 2014), as follows: The distribution of weight among phonological, orthographical, and semantic presentations is evolving in the language acquisition process. This further suggests that L2 learners with different levels of language proficiency face different challenges in the process of L2 acquisition. Although high-proficiency learners may focus on enhancing processing automaticity, those with limited L2 proficiency may consider processing accuracy to be of paramount importance.

Conclusions and limitations

The current study indicated that word statistical properties, that is, reading consistency and lexical frequency, significantly affect the L2 oral reading performance of Japanese kanji words. It provides evidence that statistical learning is the foundation of L2 word phonological processing, where words with weaker statistical property negatively affect processing accuracy and speed. Further, the impact of word statistical property is modulated by different levels of L2 proficiency. Specifically, the reading accuracy of less proficient L2 learners is highly vulnerable to the impact of the quasi-regular structure when words are in low lexical frequency.

These results can be interpreted by not only the achievement gap in L2 proficiency but also the interaction between semantic memory and word statistical properties in Japanese phonological processing.

Several limitations in this study warrant further exploration. First, the features of L1 knowledge need to be further clarified in understanding the difficulties of Japanese Kanji word processing. Although the confounding effects of the two L1-transfer factors were not observed in the present study, the existence of orthographical parallels to Chinese can potentially activate both Japanese and Chinese semantic representations. For instance, both Japanese kanji and Chinese characters can be read using phonemic or semantic radicals. In the Chinese character 錶 (watch), the semantic radical 金 (gold) suggests its meaning, while the phonemic radical 表 indicates pronunciation, as in other characters like 錶, pronounced “biao.” L2 learners may practice this orthographic-to-phonology in radical parallels for Japanese kanji which leads to reading errors, although not due to any lack of word semantic knowledge. Therefore, it is necessary to ascertain whether the lack of vocabulary (semantic knowledge) is the cause of an issue or whether the incomplete knowledge of the word phonological presentation is responsible for erroneous responses in future work. Second, while L2 proficiency was assessed through JLPT scores and academic background, some participants lacked formal test results, and future studies could benefit from using standardized online proficiency tests. Furthermore, the word materials used in the two experiments may have exceeded the reading proficiency level of less proficient learners, suggesting a need to adjust difficulty levels in future studies.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0142716425000128>

Replication package. All data, supplementary materials, and analysis code for this study are available at the first author's Open Science Framework repository (<https://osf.io/9r8av/>).

Acknowledgements. This research was supported in part by the Research Development Program “ISHIZUE” from Kyoto University to SS, and by a research grant from Kyoto University Kyoyukai, Faculty of Education Alumni Association, awarded to NK. Support for English proofreading was provided to NK by the Global Education Office, Graduate School of Education, Kyoto University. This study was not preregistered.

Notes

1 Whether the access to the phonological presentation of a kanji is directly mediated by its lexical semantics remains unclear (Hino et al., 2017; Wydell et al., 1993). However, reading knowledge from the whole-word level is not exclusive to the activation of word meaning. Fushimi et al. (1999) used non-word kanji items (e.g., 神間, meaningless but made up of two single kanji characters and still pronounceable) as stimuli in their second experiment and clarified that the reading of such items was based on a whole-word (non-word) level phonological processing.

2 The categorization of consistent and inconsistent Japanese kanji words did not originate with Fushimi et al. (1999). Wydell et al. (1995) first introduced this classification based on whether each constituent character had a single on-reading pronunciation. As the categorization of reading consistency in Wydell et al. (1995) did not prove robust impact in oral reading tasks, the present study adopts the measure of reading consistency proposed by Fushimi et al. (1999), which is grounded in the statistical property of kanji pronunciation.

- 3 The consistency effect of Japanese kanji word reading is also known as the phonological-orthographic consistency effect (Hino et al., 2011; Hino et al., 2017).
- 4 Although Standard Mandarin is the official language of natives in mainland China, dialects may potentially play a part in Japanese word phonological processing. However, as earlier studies have rarely reported dialects' impact on the L2 acquisition of Japanese kanji reading and it is beyond the scope of this study, dialects are not considered in this study. Although Standard Mandarin, known as Simplified Chinese, is used exclusively in mainland China, the fundamental semantics of Mandarin allow for an understanding of Traditional Mandarin.
- 5 The survey sheet is present in the OSF file *Recruitment of the Participants* (<https://osf.io/9r8av/>).
- 6 For further information on the word frequency list, see <https://clrd.ninjal.ac.jp/bccwj/bcc-chu.html>.
- 7 Although orthographical, phonological, and semantic similarities between Japanese kanji and Chinese characters are beyond this study's scope, their influence cannot be ignored. For instance, several *on*-reading Japanese kanji words were reported to have advantages in the reading performance of L1-Chinese learners (Kayamoto, 2000, 2002; Matsumi et al., 2014).
- 8 In the mixed-effects model analyses, contrasts for categorical variables (reading consistency and lexical frequency) were set to sum contrasts. This choice aligns with the focus on testing main effects and interactions in an orthogonal manner to prevent redundancy in variance estimation and ensure clear interpretation of interactions (Schad et al., 2020).
- 9 Irregularities refer to the program running errors that are occasionally activated in Psychopy 3.0, such as the skipping of word stimuli, or the duration of word stimuli is less than 5,000 ms or more than this value owing to network transmission lags in ZOOM broadcasts. The occurrence of such irregular situations is generally unpredictable and unavoidable in online experiments. Additionally, as reading accuracy focuses on the correctness of word pronunciation, incorrect responses, and no-response errors were included in the analysis and coded as reading errors.
- 10 By definition, no LARC error is supposed to be observed in consistent words as there is no alternative pronunciation of the constituent kanji within the words. However, one LARC error was observed in the LF consistent word group, in which the pronunciation of the first kanji character had a second legitimate pronunciation when read as part of people's names. The other two LARC errors were observed for kanji characters with rarely seen pronunciations, that is, the kanji character 軍 in the word 軍備/guN-bi/ (ministry force) was pronounced as /kuN/ and the character 材 in the word 製材/sei-zai/ (sawing) was pronounced as /sai/. Such a case was observed in the oral word reading performance by the semantic dementia patient NK (Patterson et al., 1995, p. 162), such that ... she even produced one such error to a Consistent word; if these words are supposed to have no alternative readings of their components, how can this be ... one of the characters in this word does have a kun-reading. The fact that this alternative reading does not appear in the very standard dictionary that we had consulted shows how rare it is, but NK clearly knew the alternative reading of this character and used it in this misreading. Hence, this study addressed the LARC error in consistent words in an identical fashion.
- 11 Owing to the spread of the coronavirus disease 2019 epidemic in mainland China, Experiment 2 was conducted online. To ensure an adequate number of participants with intermediate L2 proficiency, Japanese as a foreign language learners were recruited from different universities in mainland China.
- 12 For further information, see <https://www.intersc.tsukuba.ac.jp/~kyoten/chuta/>.
- 13 In the oral reading tasks, self-corrections in responses were occasionally observed for the word stimuli. While such responses were ultimately considered correct, the types of errors were recorded if they occurred in trials with self-corrected responses. As a result, in some cases, the total error rates across different types exceeded 1.

References

- Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory and Cognition*, 10(6), 565–575. <https://doi.org/10.3758/BF03202439>
- Andrews, S., Woollams, A., & Bond, R. (2005). Spelling-sound typicality only affects words with digraphs: Further qualifications to the generality of the regularity effect on word naming. *Journal of Memory and Language*, 53(4), 567–593. <https://doi.org/10.1016/j.jml.2005.04.002>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using *lme4*. *arXiv preprint arXiv:1406.5823*. <https://doi.org/10.48550/arXiv.1406.5823>

- Chao, P. C., Chen, W. F., Zevin, J., & Lee, C. Y. (2021). Neural correlates of phonology-to-orthography mapping consistency effects on Chinese spoken word recognition. *Brain and Language*, **219**, 104961. <https://doi.org/10.1016/j.bandl.2021.104961>
- Chen, W. F., Chao, P. C., Chang, Y. N., Hsu, C. H., & Lee, C. Y. (2016). Effects of orthographic consistency and homophone density on Chinese spoken word recognition. *Brain and Language*, **157–158**, 51–62. <https://doi.org/10.1016/j.bandl.2016.04.005>
- Chiu, H. C. (2002a). The processing routes of Japanese kanji words for the Taiwanese learners of Japanese: The comparison of cognates and non-cognates between Japanese-Chinese. *Bulletin of the Graduate School of Education, Hiroshima University, Part II. Arts and Science Education*, **51**, 357–365. <https://doi.org/10.15027/16210>
- Chiu, H. C. (2002b). Processing orthography and phonology in semantic decision tasks. *Japanese Journal of Educational Psychology*, **50**(4), 412–420. https://doi.org/10.5926/jjep1953.50.4_412
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed- processing approaches. *Psychological Review*, **100**(4), 589–608. <https://psycnet.apa.org/doi/10.1037/0033-295X.100.4.589>. <https://doi.org/10.1037/0033-295X.100.4.589>
- Coltheart, M., Tree, J. J., & Saunders, S. J. (2010). Computational modeling of reading in semantic dementia: comment on Woollams, Lambon Ralph, Plaut, and Patterson (2007). *Psychological Review*, **117**(1), 256–271. <https://doi.org/10.1037/a0015948>
- Crisp, J., & Lambon Ralph, M. A. (2006). Unlocking the nature of the phonological-deep dyslexia continuum: the keys to reading aloud are in phonology and semantics. *Journal of Cognitive Neuroscience*, **18**(3), 348–362. <https://doi.org/10.1162/089892906775990543>
- Cumming, G., & Finch, S. (2005). Inference by eye: confidence intervals and how to read pictures of data. *American Psychologist*, **60**(2), 170–180. <https://doi.org/10.1037/0003-066X.60.2.170>
- Fei, X. (2015). The process of learning Japanese kanji (Chinese character) words in Chinese-native learners of the Japanese language: effects of orthographical and phonological similarities between the Chinese and the Japanese languages. *Theory and Research for Developing Learning Systems*, **1**, 57–69. <https://ir.lib.hiroshima-u.ac.jp/00039417>
- Feldman, N. H., Goldwater, S., Dupoux, E., & Schatz, T. (2022). Do infants really learn phonetic categories? *Open Mind*, **5**, 113–131. https://doi.org/10.1162/opmi_a_00046
- Fisher, C., Klingler, S. L., & Song, H. -J. (2006). What does syntax say about space? 2-year-olds use sentence structure to learn new prepositions. *Cognition*, **101**(1), B19–B29. <https://doi.org/10.1016/j.cognition.2005.10.002>
- Fox, J., & Weisberg, S. (2023). *car: Companion to Applied Regression*. R package version 3.1-4. Retrieved from <https://CRAN.R-project.org/package=car>
- Fushimi, T., Ijuin, M., Patterson, K., & Tatsumi, I. F. (1999). Consistency, frequency, and lexicality effects in naming Japanese kanji. *Journal of Experimental Psychology: Human Perception and Performance*, **25**(2), 382–407. <https://doi.org/10.1037/0096-1523.25.2.382>
- Fushimi, T., Komori, K., Ikeda, M., Lambon Ralph, M. A., & Patterson, K. (2009). The association between semantic dementia and surface dyslexia in Japanese. *Neuropsychologia*, **47**(4), 1061–1068. <https://doi.org/10.1016/j.neuropsychologia.2008.12.030>
- Fushimi, T., Komori, K., Ikeda, M., Patterson, K., Ijuin, M., & Tanabe, H. (2003). Surface dyslexia in a Japanese patient with semantic dementia: Evidence for similarity-based orthography-to-phonology translation. *Neuropsychologia*, **41**(12), 1644–1658. [https://doi.org/10.1016/S0028-3932\(03\)00108-8](https://doi.org/10.1016/S0028-3932(03)00108-8)
- Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, **5**(4), 674–691. <https://doi.org/10.1037/0096-1523.5.4.674>
- Halai, A. D., Woollams, A. M., & Lambon Ralph, M. A. (2018). Triangulation of language-cognitive impairments, naming errors and their neural bases post-stroke. *NeuroImage: Clinical*, **17**, 465–473. <https://doi.org/10.1016/j.nicl.2017.10.037>
- Hino, Y., Kusunose, Y., Miyamura, S., & Lupker, S. J. (2017). Phonological-orthographic consistency for Japanese words and its impact on visual and auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **43**(1), 126–146. <https://doi.org/10.1037/xhp0000281>
- Hino, Y., Miyamura, S., & Lupker, S. J. (2011). The nature of orthographic-phonological and orthographic-semantic relationships for Japanese kana and kanji words. *Behavior Research Methods*, **43**(4), 1110–1151. <https://doi.org/10.3758/s13428-011-0101-0>

- Ida, K., Yoshihara, M., Xue, J., Kusunose, Y., Sato, H., & Hino, Y. (2014). Effects of orthographic-phonological consistency and on/kun-readings in the naming of Japanese kanji words. *Japanese Journal of Cognitive Psychology*, 11(2), 117–126. <https://doi.org/10.5265/jcogpsy.11.117> (In Japanese with English abstract).
- Ijuin, M., Fushimi, T., & Tatsumi, I. F. (1996). Consistency effect in naming Japanese kanji words: Computational by neural network. *Proceedings of the Japanese Psychological Association*, 60, 609. <https://doi.org/10.11225/jcss.9.187> (In Japanese with English abstract).
- Jared, D. (1997). Spelling-sound consistency affects the naming of HF words. *Journal of Memory and Language*, 36(4), 505–529. <https://doi.org/10.1006/jmla.1997.2496>
- Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, 29(6), 687–715. [https://doi.org/10.1016/0749-596X\(90\)90044-Z](https://doi.org/10.1016/0749-596X(90)90044-Z)
- Kayamoto, Y. (2000). Processing phonological information: Recognition of Japanese characters by advanced- and superior-level native speakers of Chinese. *Japanese Journal of Educational Psychology*, 48(3), 315–322. https://doi.org/10.5926/jjep1953.48.3_315 (In Japanese).
- Kayamoto, Y. (2002). Lexical access to Japanese kanji by native speakers of Chinese: Evidence from lexical decision and naming tasks. *Japanese Journal of Educational Psychology*, 50(4), 436–445. https://doi.org/10.5926/jjep1953.50.4_436 (In Japanese).
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). *lmerTest* package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., & Jensen, S. P. (2020). *lmerTest* (Version 3.1-3) [Computer software]. <https://github.com/runehaubo/lmerTestR>
- Lee, C. Y., Tsai, J. L., Su, E. C. I., Tzeng, O. J., & Hung, D. L. (2005). Consistency, regularity, and frequency effects in naming Chinese characters. *Language and Linguistics*, 6(1), 75–107.
- Lenth, R. V. (2023). emmeans: Estimated Marginal Means, aka Least-Squares Means (R package version 1.10.5). Retrieved from <https://rvlenth.github.io/emmeans/>
- Li, M. F., Gao, X. Y., & Wu, J. T. (2020). Neighborhood effects in Chinese character recognition: going beyond phonological perspectives to explain a possible underlying mechanism. *Reading and Writing*, 33, 547–570. <https://doi.org/10.1007/s11145-019-09973-4>
- Matsumi, N., Fei, X., & Cai, F. (2014). The influence of Japanese language on the phonological processing of Chinese words in Chinese students learning Japanese as a second language: Operating the orthographical and phonological similarities of kanji-words. *Bulletin of the Graduate School of Education, Hiroshima University, Part II. Arts and Science Education*, 63, 191–198. <https://doi.org/10.15027/36647> (In Japanese with English abstract).
- Nishio, M., Iwabuchi, E., & Mizutani, S. (Eds.). (1986). *Iwanami kokugo jiten [Iwanami Japanese dictionary]* (4th ed.). Iwanami Shoten.
- Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32(2), 258–278. <https://doi.org/10.1006/jmla.1993.1014>
- Patterson, K., & Lambon Ralph, M. A. (1999). Selective disorders of reading? *Current Opinion in Neurobiology*, 9(2), 235–239. [https://doi.org/10.1016/S0959-4388\(99\)80033-6](https://doi.org/10.1016/S0959-4388(99)80033-6)
- Patterson, K., Suzuki, T., Wydell, T., & Sasanuma, S. (1995). Progressive aphasia and surface alexia in Japanese. *Neurocase*, 1(2), 155–165. <https://doi.org/10.1080/13554799508402358>
- Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods*, 51, 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, 103(1), 56–115. <https://doi.org/10.1037/0033-295X.103.1.56>
- Richtsmeyer, P. T., Gerken, L., & Ohala, D. K. (2011). Contributions of phonetic token variability and word-type frequency to phonological representations. *Journal of Child Language*, 38(5), 951–978. <https://doi.org/10.1017/S0305000910000371>
- Richtsmeyer, P. T., & Goffman, L. (2017). Perceptual statistical learning over one week in child speech production. *Journal of Communication Disorders*, 68, 70–80. <https://doi.org/10.1016/j.jcomdis.2017.06.004>

- Richtsmeier, P. T., & Goffman, L. (2023). Statistical learning of phonotactics by children can be affected by another statistical learning task. *Applied Psycholinguistics*, 44(6), 1124–1142. <https://doi.org/10.1017/S0142716423000449>
- Ripley, B., Venables, W., & Ripley, M. B. (2016). Package ‘nnet’. *R package* (Version 7.3-12). <https://cran.r-project.org/package=nnet>
- Roux, F., Armstrong, B. C., & Carreiras, M. (2017). Chronset: an automated tool for detecting speech onset. *Behavior Research Methods*, 49(5), 1864–1881. <https://doi.org/10.3758/s13428-016-0830-1>
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: a tutorial. *Journal of Memory and Language*, 110, 104038. <https://doi.org/10.1016/j.jml.2019.104038>
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523–568. <https://doi.org/10.1037/0033-295X.96.4.523>
- Singmann, H., Bolker, B., Westfall, J., Aust, F., Ben-Shachar, M. S., Højsgaard, S., ... & Love, J. (2015). Package ‘afex’. Retrieved from <https://github.com/singmann/afex>
- Tamaoka, K., Makioka, S., Sanders, S., & Verdonchot, R. G. (2017). www.kanjidatabase.com: a new interactive online database for psychological and linguistic research on Japanese kanji and their compound words. *Psychological Research*, 81, 696–708. <https://doi.org/10.1007/s00426-016-0764-3>
- 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. https://doi.org/10.1162/jocn_a_00467
- Wydell, T. N., Butterworth, B., & Patterson, K. (1995). The inconsistency of consistency effects in reading: the case of Japanese kanji. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(5), 1155–1168. <https://doi.org/10.1037/0278-7393.21.5.1155>
- Wydell, T. N., Patterson, K. E., & Humphreys, G. W. (1993). Phonologically mediated access to meaning for kanji: Is a row still a rose in Japanese kanji? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(3), 491–514. <https://doi.org/10.1037/0278-7393.19.3.491>

Cite this article: Kang, N. & Saito, S. (2025). Phonological processing of Japanese *Kanji* word by L1-Chinese learners of Japanese: verification of consistency and frequency effects. *Applied Psycholinguistics*. <https://doi.org/10.1017/S0142716425000128>