



The effect of the position of atypical character-to-sound correspondences on reading kanji words aloud: Evidence for a sublexical serially operating kanji reading process

Ami Sambai^{1,2} · Max Coltheart³ · Akira Uno^{2,4}

Published online: 5 February 2018
© Psychonomic Society, Inc. 2018

Abstract

In English, the size of the regularity effect on word reading-aloud latency decreases across position of irregularity. This has been explained by a sublexical serially operating reading mechanism. It is unclear whether sublexical serial processing occurs in reading two-character kanji words aloud. To investigate this issue, we studied how the position of atypical character-to-sound correspondences influenced reading performance. When participants read inconsistent-atypical words aloud mixed randomly with nonwords, reading latencies of words with an inconsistent-atypical correspondence in the initial position were significantly longer than words with an inconsistent-atypical correspondence in the second position. The significant difference of reading latencies for inconsistent-atypical words disappeared when inconsistent-atypical words were presented without nonwords. Moreover, reading latencies for words with an inconsistent-atypical correspondence in the first position were shorter than for words with a typical correspondence in the first position. This typicality effect was absent when the atypicality was in the second position. These position-of-atypicality effects suggest that sublexical processing of kanji occurs serially and that the phonology of two-character kanji words is generated from both a lexical parallel process and a sublexical serial process.

Keywords Sublexical serial processing · Kanji reading · Position of atypicality effect

The Japanese writing system is unique. It consists of two qualitatively different scripts: logographic kanji and two forms of syllabic kana, hiragana and katakana. The two kana scripts comprise a transparent writing system. On the other hand, in the case of kanji, the relationship between orthography and phonology is very opaque. Kanji characters were originally borrowed from Chinese characters. Most kanji characters have multiple readings. Some kanji characters appear as single-character words as well as components of multiple-character kanji words. Others never appear as single-

character words: They only appear as components of multiple-character kanji words. Eighty percent of all kanji words are two-character kanji compound words.

So far, models of word recognition and reading aloud for two-character kanji words have been investigated empirically and computationally. There seems to be agreement that the phonology of a kanji word is generated by both lexical and sublexical processing. With regard to the sublexical process, all models assume that each constituent character of a word is converted into its corresponding sound (e.g., Fushimi, Ijuin, Patterson, & Tatsumi, 1999; Ijuin, Fushimi, Patterson, & Tatsumi, 1999; Miwa, Libben, Dijkstra, & Baayen, 2014; Saito, Masuda, & Kawakami, 1998). In addition, some have argued that phonological information of both whole characters and their component radical(s) are automatically activated (Saito et al., 1998), whereas others have denied this, for example, “Kanji characters cannot be decomposed into orthographic elements corresponding to phonology at anything below the level of the whole character” (Fushimi et al., 1999, p. 383).

When English words are read aloud, reading latencies are longer when the words are irregular in grapheme–phoneme

✉ Ami Sambai
sambai@cc.osaka-kyoiku.ac.jp

¹ Osaka Kyoiku University, Osaka, Japan

² LD/Dyslexia Centre, Chiba, Japan

³ Macquarie University, Sydney, New South Wales, Australia

⁴ University of Tsukuba, Ibaraki, Japan

correspondence than when they are regular (e.g., Seidenberg, Waters, Barnes, & Tanenhaus, 1984). This phenomenon is known as the regularity effect. Furthermore, the regularity effect is larger when the irregular grapheme-to-phoneme correspondence is in the initial position (e.g., *chef*) than when it is in later positions (e.g., *deaf*; e.g., Coltheart & Rastle, 1994; Cortese, 1998; Rastle & Coltheart, 1999; Roberts, Rastle, Coltheart, & Besner, 2003). This interaction is called the position of irregularity effect.¹

The interpretation of the position of irregularity effect as a serial effect has been developed by the dual-route approach (e.g., Coltheart, Rastle, Perry, Ziegler, & Langdon, 2001). The dual-route approach assumes that the phonology of a word is generated by a lexical dictionary look-up procedure (which acts in parallel) and a sublexical rule-based procedure (which acts serially left to right). When people read irregular words aloud, they need to solve the conflict between the correct phonology derived from the lexical route and the incorrect phonology derived from the sublexical route. As a result, irregular words are read more slowly than regular words. Moreover when irregular words have their irregular grapheme-to-phoneme correspondence in the initial position (e.g., *chef*), the sublexical route will produce the incorrect phoneme for the initial grapheme before the correct output from the lexical route is activated sufficiently. In contrast, when irregular words have their irregular grapheme-to-phoneme correspondence in later positions, the correct output from the lexical route will be activated sufficiently before the wrong output from the sublexical route for the irregular correspondence is activated. Therefore, the conflict between the two routes will be stronger for words with irregular grapheme-to-phoneme correspondence in the initial position than for words with irregular grapheme-to-phoneme correspondence in later positions. Hence, the size of the regularity effect will vary with position of irregularity.

For the kanji writing system, character-to-sound mappings are not rule based. There are no regular character-to-sound correspondences exactly equivalent to those in English words (Wydell, Butterworth, & Patterson, 1995), so the concept of regularity does not apply. Instead, the consistency effect is used as a benchmark effect showing that sublexical processing is occurring during reading kanji words aloud, in place of a regularity effect.

In Wydell et al.'s (1995) study, consistency was defined in term of ON-reading and KUN-reading. ON-reading is a pronunciation derived from that of an

Chinese character, for example, /shin/ for the kanji character “親”. KUN-reading is a pronunciation derived from that of an original Japanese word that has the same meaning as the Chinese character, for example, /oya/ (*parent*, in English) for the same kanji character (see Wydell et al., 1995, for details). For characters with multiple readings, the appropriate pronunciation is assumed to be determined by within-word context, that is, by another constituent character of the two-character kanji word. For example, the kanji character “親” is seen in multiple kanji compound words. When it appears in two-character kanji words such as “親” /oyakata/ (*foreman*, in English) and “父親” /chichioya/ (*father*, in English), it is pronounced as its KUN-reading /oya/. On the other hand, it is pronounced as its ON-reading /shin/ when it appears in other two-character kanji words such as “親□” /shinsetsu/ (*kind*, in English) and “両親” /ryoshin./ (*parents*, in English). Generally the ON pronunciation is the more typical—that is, multicharacter words most often have ON pronunciations for their component characters. Wydell et al. made use of these characteristics to define consistency. A two-character kanji word was considered *consistent* if each constituent character has only an ON-reading but no KUN-reading. A two-character kanji word was considered *Intermediate 1* if either the first or the second character of the word, when it appears as a single character, has a KUN-reading as well as an ON-reading, though its KUN-reading does not apply to any two-character words. A two-character kanji word was considered *Intermediate 2* if both the first and the second characters of the word have KUN-readings when they appear as single characters, though these KUN-readings do not occur in other two-character words. A two-character kanji word was considered *inconsistent* if either one or both component characters of the word have KUN-readings, and these KUN-readings are appropriate to other words containing the characters. Wydell et al. (1995) found that reading latencies did not differ across these four conditions (i.e., there were no consistency effects).

In contrast, Fushimi et al. (1999) did report a significant consistency effect in reading two-character kanji words. They introduced the idea of graded consistency, that is, the degree of character-to-sound correspondences that is calculated in a corpus of two-character kanji words statistically, in place of the classification based on ON-readings and KUN-readings. In their classification, character-to-sound correspondences fall into three types based on the proportion of orthographic friends to orthographic neighbors. When a character has the same pronunciation across all words containing that character at the same position, the character was classified as *consistent*. When a character has multiple pronunciations but its pronunciation is the most common in words containing that character (i.e., the pronunciation of a character has more friends than

¹ Cortese and Simpson (2000) argued that at least some part of the regularity effect is due to body-rime inconsistency rather grapheme-phoneme irregularity. This point, however, is not relevant to the effect of position of regularity, which is the concern of our study.

any other pronunciations of the character), the character was classified as *inconsistent-typical*. Finally, when a character has multiple pronunciations and its pronunciation is not the most common in words containing that character (i.e., the pronunciation of a character does not have more friends than other pronunciations of the character), the character is classified as *inconsistent-atypical*. If a word consisted of two characters that were both of the consistent type, the word was referred to as a consistent word. If a word consisted of two characters that were both of the inconsistent-typical type, the word was referred to as an inconsistent-typical word. If a word had at least one character that was the inconsistent-atypical type, the word was referred to as an inconsistent-atypical word. Reading-aloud latencies showed a significant interaction between consistency and frequency. Inconsistent-atypical words produced longer latencies than both consistent words and inconsistent-typical words, but only for low-frequency words. Consistent words and inconsistent-typical words did not differ in reading latency. The same pattern as observed with low-frequency kanji words was also observed in reading aloud of two-character kanji nonwords. A two-character kanji nonword is one where each character corresponds to a real word and so has a pronunciation, but the combination of these two characters does not correspond to any real word (e.g., a nonword “信別” /shiNbetsu/ was generated by combining the first character of a word “信任[confidence, in English]”/shiNniN/ and the second character of a word “区别 (distinction, in English)” /kubetsu/).

The difference between the results of Wydell et al. (1995) and Fushimi et al. (1999) would seem to be due to different definitions of consistency. According to the definition of Fushimi et al., Wydell et al.’s consistent and inconsistent words were mainly consistent and inconsistent-typical words, which is why there was no significant consistency effect. Fushimi et al.’s significant consistency effects are taken as a benchmark effect showing the sublexical reading procedure at the character level: That is, with a two-character word, not only is the phonology of the whole word activated but also is the phonology of its individual character components. Given that reading latencies vary across the position of an atypical character-to-sound correspondence, this reading phenomenon, that is, an effect of the position of atypicality (corresponding to a position of irregularity effect in English word reading), could be strong evidence supporting for the existence of a sublexical serially operating reading process during reading a two-character kanji word aloud.

The aim of this study was to determine whether kanji character-to-sound conversion is applied to the individual characters of two-character kanji words serially or in parallel. We pursued this aim by investigating whether position of an inconsistent-atypical correspondence influenced reading performance significantly.

Experiment 1

Method

Participants

Forty-six undergraduate or postgraduate students at the University of Tsukuba participated. All were native speakers of Japanese and had normal or corrected-to-normal vision and no reading impairment.

Stimuli

Seventy-eight two-character kanji words were used as targets (see [Appendix](#)). They were selected from the NTT Psycholinguistic Database (Amano & Kondo, 1999). All were middle-frequency or low-frequency words, with logarithmic frequencies at least one standard deviation below the mean of all two-character kanji words.

As described in the Introduction, Fushimi et al. (1999) classified character-to-sound correspondences for each character position of two-character kanji words into *consistent* (i.e., a character has the same pronunciation across all words containing that character in that position), *inconsistent-typical* (i.e., a character has multiple pronunciations, but its pronunciation is the most common in words containing that character in that position), or *inconsistent-atypical* (i.e., a character has multiple pronunciations, and its pronunciation is not the most common in words containing that character in that position). In our study, consistent or inconsistent-typical readings were referred to as *typical* and inconsistent-atypical readings were referred to as *atypical*. As in Fushimi et al.’s study, our classification of character-to-sound correspondences was made for first and second characters separately, following the proportion of friends to neighbors that were computed for each character position.

Stimuli in our study fell into three conditions based on the position of atypical readings. The first condition consisted of words with typical readings in the first position and atypical readings in the second position (typical-atypical condition). The second condition consisted of words with atypical readings in the first position and typical readings in the second position (atypical-typical condition). The third condition consisted of words with atypical readings in both positions (atypical-atypical condition). Each condition had 26 stimuli.

To match degree of consistency of character-to-sound correspondences (consistency value) between the initial and second positions, we computed consistency values for first and second characters separately in two-character kanji words. Consistency values were given by dividing the number of words with the same character and the same pronunciation of that character (i.e., number of friends) by the number of words possessing that character regardless of its sound (i.e.,

number of neighbors), following the definition of consistency value in Fushimi et al. (1999). Neighborhood size and consistency value were calculated for each character position, using the corpus of all 15,256 two-character kanji words seen in the NTT Psycholinguistic Database. In our stimuli, mean neighborhood size was controlled in each position across all of the three word conditions, as was mean consistency value of typical readings and atypical readings.

Mean auditory familiarity, orthographic familiarity, auditory imageability, orthographic imageability, and log frequency were also controlled across all conditions, as was the year of compulsory education when the character was introduced to children for the first time (age of acquisition). These word attributes were taken from the NTT Psycholinguistic Database (Amano & Kondo, 1999). Table 1 shows means and standard deviations of word attributes. To avoid confounding effects of initial morae on measurements of reading latencies, initial morae were matched across the conditions.

We added 100 two-character kanji nonwords as fillers. All were generated by combining characters that were not used in the word stimuli. Although these are nonwords in the sense that the two kanji characters in each nonword never occur together, they can still be read aloud because each individual character in the pair has a pronunciation. Consistency (i.e., consistent or inconsistent character–sound relationships) in the initial character position was crossed with consistency in the last character position, which generated a total of four types consisting of equal number of stimuli. All inconsistent characters had their inconsistent-typical readings with consistency values higher than 0.5 and lower than 0.87 (mean = 0.69

± 0.07). This means that any inconsistent readings had more friends than enemies.

In addition to the experimental items, 12 words and 18 nonwords were used as practice items.

Apparatus

Stimulus presentation and data recording were controlled by E-Prime 2.0 (Psychology Software Tools, Inc, USA) on the 12.5-inch monitor of a laptop computer (Intel^R Celeron^R M CPU, 1.2 GHz). Reading latencies were obtained using a voice key connected to the laptop computer via a headset microphone fitted to each participant.

Procedure

Participants were tested individually. They were seated approximately 50 cm in front of the monitor. Stimuli were presented in MS Mincho (font size 18) on the screen. Participants were instructed that each stimulus would be presented after a fixation cross, and that their task was to read aloud each stimulus as quickly and accurately as possible.

The 178 experimental trials were conducted after 30 practice trials. Each trial started with the presentation of a fixation cross for 500 ms. Then, the target stimulus was presented. It remained on the monitor screen until the participant had responded. The intertrial interval was 1,000 ms. Order of presentation of stimuli was randomized across participants. There were breaks every 35 trials.

Table 1 The means and standard deviations of word attributes in all stimuli

	Atypical-atypical words		Atypical-typical words		Typical-atypical words		Typical-typical words	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mora length	3.69	0.72	3.54	0.63	3.65	0.62	3.62	0.49
Auditory familiarity	5.28	0.36	5.28	0.33	5.22	0.46	5.39	0.35
Orthographic familiarity	5.37	0.31	5.27	0.39	5.17	0.37	5.28	0.35
Auditory imageability	4.27	0.36	4.26	0.51	4.22	0.36	4.32	0.49
Orthographic imageability	4.32	0.44	4.26	0.58	4.17	0.43	4.20	0.54
Log frequency	5.28	1.19	5.34	1.29	5.10	1.26	5.40	1.21
Age of acquisition in the first position	2.62	1.04	2.42	1.50	3.19	1.78	3.12	1.34
Age of acquisition in the second position	2.77	1.55	2.88	1.60	2.35	1.17	3.04	1.16
Neighborhood size in the first position	22.81	11.94	22.58	12.95	18.62	14.87	23.96	15.95
Neighborhood size in the second position	33.77	18.89	29.96	21.25	32.81	17.39	23.31	13.29
Number of friends in the first position	4.54	4.13	5.58	3.86	16.04	11.22	21.46	13.24
Number of friends in the second position	8.00	6.29	25.62	18.58	6.77	4.54	20.96	12.40
Consistency value in the first position	0.21	0.13	0.25	0.11	0.89	0.12	0.92	0.09
Consistency value in the second position	0.25	0.13	0.85	0.13	0.22	0.12	0.89	0.10

Note. SD = standard deviation. Typical-typical words were used in only Experiment 3

Results

Reading accuracies of two typical-atypical words (庭石/niwaishi/ and 愛想/aiso/) and two atypical-typical words (角地/kadochi/ and 遠縁/to:en/) were less than 60%. Therefore, these four words were excluded from the statistical analyses. In addition, the words matching them on the first mora in other conditions also were removed. Thus, there were 22 stimuli available to statistical analyses per word condition. In these remaining stimuli, word attributes were controlled across all conditions: mean auditory familiarity, orthographic familiarity, auditory imageability, orthographic imageability, log frequency, age of acquisition in each character position, and *N* size in each character position were not statistically different across all conditions. Moreover, consistency value for atypical readings and consistency value for typical readings were also matched between the initial and second position. Data from three participants were excluded because they made more than 30% errors for atypical-atypical words.

Reading latencies (RTs)

Reading latencies available to statistical analyses were determined by the following procedure. Any reading latencies of target items that were not read accurately, or that were self-corrected, were excluded. In addition, any reading latencies that were less or more than 2.5 standard deviations away from the mean of each condition for each participant were excluded. Reading latencies of nonword fillers were also trimmed by the same procedure. Table 2 shows the mean and standard deviations of reading latencies and error rates (calculated from subject means). A total of 12.5% reading latencies was excluded. This overall exclusion rate was relatively high because the rate of misreadings was relatively high (6.9%).

Logarithmic reading latencies were analyzed by mixed-effects modeling (Baayen, 2008). Subjects and items were treated as random effect factors. Word type (typical-atypical, atypical-typical, atypical-atypical) was treated as a fixed-effect factor. Trial number and previous RTs (i.e., RT for the preceding trial) were included as a covariate.

When the mixed-effects Model 1 (see Table 3) including two interactions (i.e., a Word Type × Previous RTs interaction and a Word Type × Trial Number interaction) and main effects

Table 3 The mixed-effect models used for analysis in Experiments 1 and 2

	Model 1	Model 2
Word Type × Previous RTs	○	
Word Type × Trial Number	○	
Word type	○	○
Previous RTs	○	○
Trial number	○	○

Note. Each model contained interactions or main effects represented by a circle

of three factors (i.e., word type, previous RTs, and trial number) compared with a mixed-effects model removing either two-way interaction from Model 1, model fit did not improve significantly (Word Type × Previous RTs: $\chi^2 = 0.024, p > .1$; Word Type × Trial Number: $\chi^2 = 0.36, p > .1$). This means that the word-type effect was not influenced by previous RTs and trial number. Using the mixed-effects Model 2 that contained only main effects (see Table 3), the word-type effect was significant. Typical-atypical words were read significantly faster than both atypical-typical words ($t = 2.31, p < .05, \beta = 7.97E-02, SE = 3.45E-02$) and atypical-atypical words ($t = 2.64, p < .01, \beta = 9.12E-02, SE = 3.45E-02$). There was no difference between atypical-typical words and atypical-atypical words ($t = 0.33, p > .1, \beta = 1.15E-02, SE = 3.45E-02$). For covariates, both the effect of trial number and the effect of previous RTs were significant (trial number: $t = 2.19, p < .05, \beta = 1.79E-04, SE = 8.16E-05$; previous RTs: $t = 3.19, p < .01, \beta = 1.88E-05, SE = 5.89E-06$).

A Bayesian one-way repeated-measures ANOVA was conducted for the subject data reported in Table 2, using JASP. Setting the default scale *r* on fixed-effect size of 0.5, for a main effect of word type, BF_{10} (i.e., BF in favor of the alternative) was $1.128E+7$. This result suggests strong evidence for H_1 (Wagenmakers et al., 2017). Bayesian paired-sample *t* tests were conducted as post hoc comparison. Setting the default scale *r* on effect size of 0.5, for the atypical-atypical words versus typical-atypical words comparison, BF_{10} was 369723.93; for the atypical-typical words versus typical-atypical words comparison, BF_{10} was 54805.74; for the atypical-atypical words versus atypical-typical words comparison, BF_{10} was 0.25. Following Wagenmakers et al. (2017),

Table 2 The means and standard deviations of reading latencies and error rates in Experiment 1

		Atypical-atypical words	Atypical-typical words	Typical-atypical words	Nonword fillers
RTs (ms)	Mean	907.86	901.86	823.97	1,146.61
	<i>SD</i>	153.99	171.21	139.05	414.20
Error rates (%)	Mean	7.5	6.1	7.2	7.0
	<i>SD</i>	26.0	24.0	26.0	26.0

Note. *SD* = standard deviation

we conclude that two-character kanji word reading is faster when the first character is typical than when it is not (atypical-atypical is slower than typical-atypical, and atypical-typical was slower than typical-atypical). We also conclude that typicality of the second character has no effect since we are able to assert the null hypothesis that atypical-atypical words do not differ in RT from atypical-typical words.

Error rates

Error data were analyzed by logistic regression, with word type as a predictor. There was no significant word type effect (typical-atypical vs. atypical-typical: $z = 0.92, p > .1$, $\beta = 0.17, SE = 0.18$; typical-atypical vs. atypical-atypical: $z = 0.26, p > .1$, $\beta = 0.047, SE = 0.18$; atypical-typical vs. atypical-atypical: $z = 1.18, p > .1$, $\beta = 0.22, SE = 0.18$).

Discussion

The significant effect of the position of atypical character-to-sound correspondences was observed in analysis of reading latencies rather than error rates. Reading latencies were significantly longer when an inconsistent-atypical character-to-sound correspondence lay in the initial position of a kanji word (i.e., atypical-typical and atypical-atypical conditions) than when an inconsistent-atypical character-to-sound correspondence lay in the second position of a kanji word (i.e., typical-atypical condition). Our results were not due to any confound with character-to-sound consistency because the degree of character-to-sound consistency was controlled for both typical and atypical readings separately across all conditions.

There has been a main challenge to the interpretation of the position of irregularity effect as a serial effect. Kawamoto, Kello, Jones, and Bame (1998) argued that if a stimulus list included words beginning with plosives and nonplosives, a position of irregularity effect can arise. This is because a regularity effect in the initial position is obtained irrespective of initial plosivity, whereas a regularity effect in the second position interacts with initial plosivity (i.e., a significant regularity effect is observed only for words beginning with plosives). Our stimuli included words beginning with plosive phonemes and words beginning with nonplosive phonemes. The factor of initial-phoneme plosivity was added as a fixed factor and subsequent analyses carried out.

Firstly, we investigated interactions of fixed factors (word type and initial-phoneme plosivity) with covariates (trial number and previous RTs). When the mixed-effects Model 1 (see Table 4) was compared with a mixed-effects model removing an interaction with either covariate from Model 1, model fit did not improve significantly (Word Type \times Initial-Phoneme Plosivity \times Previous RTs: $\chi^2 = 1.84, p > .1$; Word Type \times Initial-Phoneme Plosivity \times Trial Number: $\chi^2 = 0.094, p > .1$; Word Type \times Previous RTs: $\chi^2 = 0.004, p > .1$; Initial-

Phoneme Plosivity \times Previous RTs: $\chi^2 = 1.17, p > .1$; Word Type \times Trial Number: $\chi^2 = 0.34, p > .1$; Initial-Phoneme Plosivity \times Trial Number: $\chi^2 = 0.27, p > .1$). This means that the investigated interactions were not significant. Then, the mixed-effects Model 2 (see Table 4) was constructed by excluding the nonsignificant interactions from Model 1. When Model 2 was compared with the mixed-effects Model 3 that contained only main effects (see Table 4), model fit did not improve significantly ($\chi^2 = 1.02, p > .1$). This means that there was no significant interaction between word type and initial-phoneme plosivity. Using Model 3, a word-type effect was significant: Typical-atypical words were read significantly faster than both atypical-typical words ($t = 2.34, p < .05$, $\beta = 7.95E-02, SE = 3.40E-02$) and atypical-atypical words ($t = 2.67, p < .01$, $\beta = 9.11E-02, SE = 3.41E-02$) but there was no difference between atypical-typical words and atypical-atypical words ($t = 0.34, p > .1$, $\beta = 1.15E-02, SE = 3.40E-02$). The initial-phoneme plosivity was not significant ($t = 1.69, p = .091$, $\beta = 5.03E-02, SE = 2.98E-02$). The effects of covariates were significant (trial number: $t = 2.20, p < .05$, $\beta = 1.79E-04, SE = 8.16E-05$; previous RTs: $t = 3.18, p < .01$, $\beta = 1.87E-05, SE = 5.89E-06$). In sum, initial-phoneme plosivity did not influence the results of Experiment 1 significantly.

All the words used in Experiment 1 were inconsistent-atypical words, which correspond to irregular words in English. They can be read aloud accurately only by the lexical reading procedure. If participants had used just the lexical reading procedure, reading latencies in the typical-atypical condition would not have differed from the other two conditions in Experiment 1. This is because the lexical reading procedure derives a pronunciation from a written word in parallel. Therefore it is likely that a sublexical reading procedure was involved in reading our kanji word stimuli. More importantly, it appears that sublexical reading processing occurs serially for the kanji writing system. If this had not been so, we would not have obtained a difference in RTs between the atypical-typical condition and the typical-atypical condition.

Our reading list had more nonword fillers than words. According to a strategic control hypothesis in English, the sublexical reading procedure is emphasized when participants read nonwords, and the lexical reading procedure is emphasized when participants read irregular words (e.g., Coltheart, 1978; Cummine, Amyotte, Pancheshen, & Chouinard, 2011; Monsell, Patterson, Graham, Hughes, & Milroy, 1992; Rastle & Coltheart, 1999; Reynolds & Besner, 2008; Zevin & Balota, 2000). If the significant position of atypicality effect in Experiment 1 was caused by the sublexical reading procedure and if nonword fillers in Experiment 1 emphasized the sublexical reading procedure, a position of atypicality effect should be reduced or even absent when there are no nonwords were present in the experiment, especially when all of the

Table 4 The mixed-effect models used for additional analysis in Experiments 1 and 2

	Model 1	Model 2	Model 3
Word Type × Initial-Phoneme Plosivity × Previous RTs	○		
Word Type × Initial-Phoneme Plosivity × Trial Number	○		
Word Type × Previous RTs	○		
Initial-Phoneme Plosivity × Previous RTs	○		
Word Type × Trial Number	○		
Initial-Phoneme Plosivity × Trial Number	○		
Word Type × Initial-Phoneme Plosivity	○	○	
Word type	○	○	○
Initial-phoneme plosivity	○	○	○
Previous RTs	○	○	○
Trial number	○	○	○

Note. Each model contained interactions or main effects represented by a circle

words used were inconsistent-atypical words. To investigate this, we carried out Experiment 2, in which all the stimuli to be read were inconsistent-atypical words; there were no nonwords.

Experiment 2

Method

Participants

Twenty-two undergraduate or postgraduate students at University of Tsukuba participated. All were native speakers of Japanese and had normal or corrected-to-normal vision and no reading impairment.

Stimuli

In contrast to Experiment 1, only words were used as stimuli. Stimuli were the same 78 two-character kanji inconsistent-atypical words as used in Experiment 1

Apparatus

The apparatus was the same as in Experiment 1.

Procedure

The procedure was the same as in Experiment 1. There were breaks every 26 trials.

Results

Data from one participant was removed from statistical analyses because the participant showed more than 30% errors for

atypical-typical words. The following analyses were conducted for just the stimuli available to statistical analyses in Experiment 1 (i.e., 22 stimuli per word condition).

Reading latencies (RTs)

Reading latencies available to statistical analyses were determined by the same procedure as in Experiment 1. Table 5 shows the mean and standard deviation of reading latencies and error rates (calculated from subject means). A total of 9.6% reading latencies was excluded. This overall exclusion rate was relatively high because the rate of misreadings was relatively high (4.0%).

Logarithmic reading latencies were analyzed by mixed-effects modeling in the same way as analyses in Experiment 1. When Model 1 (see Table 3) was compared with a mixed-effects model removing both two-way interactions from Model 1, model fit did not improve significantly (Word Type × Previous RTs: $\chi^2 = 1.53$, $p > .1$; Word Type × Trial Number: $\chi^2 = 0.093$, $p > .1$). This means that the word-type effect was not influenced by previous RTs and trial number. Using Model 2 that contained only main effects (see Table 3), there was no significant effect of word type (typical-atypical vs. atypical-typical: $t=0.37$, $p > .1$, $\beta = 1.51E-02$, $SE = 4.05E-02$; typical-atypical vs. atypical-atypical: $t = 0.44$, $p > .1$, $\beta = 1.79E-02$, $SE = 4.05E-02$; atypical-typical vs. atypical-atypical: $t = 0.07$, $p > .1$, $\beta = 2.83E-03$, $SE = 4.04E-02$). Neither trial number effect nor previous RTs effect were significant (trial number: $t = 1.20$, $p > .1$, $\beta = 3.51E-04$, $SE = 2.93E-04$; previous RTs: $t = 0.74$, $p > 0.1$, $\beta = 9.18E-06$, $SE = 1.24E-05$).

The factor of initial-phoneme plosivity was added as a fixed factor and used in further analyses. When Model 1 (see Table 4) was compared with a mixed-effects model removing an interaction with either covariate from Model 1, model fit did not improve significantly (Word Type × Initial-

Table 5 The means and standard deviations of reading latencies and error rates in Experiment 2

		Atypical-atypical words	Atypical-typical words	Typical-atypical words
RTs (ms)	Mean	846.74	838.42	821.27
	SD	161.88	164.68	153.29
Error rates (%)	Mean	3.7	4.0	4.3
	SD	19.0	19.0	20.0

Note. *SD* = standard deviation

Phoneme Plosivity \times Previous RTs: $\chi^2 = 3.14$, $p > .1$; Word Type \times Initial-Phoneme Plosivity \times Trial Number: $\chi^2 = 0.44$, $p > .1$; Word Type \times Previous RTs: $\chi^2 = 1.52$, $p > .1$; Initial-Phoneme Plosivity \times Previous RTs: $\chi^2 = 0.073$, $p > .1$; Word Type \times Trial Number: $\chi^2 = 0.07$, $p > .1$; Initial-Phoneme Plosivity \times Trial Number: $\chi^2 = 0.087$, $p > .1$). This means that covariates did not influence effects of fixed factors. When Model 2 (see Table 4) was compared with Model 3 (see Table 4), model fit did not improve significantly ($\chi^2 = 0.038$, $p > .1$). This means that there was no significant interaction between word type and initial-phoneme plosivity. Using Model 3, there was no significant effect of word type (typical-atypical vs. atypical-typical: $t = 0.37$, $p > .1$, $\beta = 1.49\text{E-}02$, $SE = 3.97\text{E-}02$; typical-atypical vs. atypical-atypical: $t = 0.45$, $p > .1$, $\beta = 1.79\text{E-}02$, $SE = 3.97\text{E-}02$; atypical-typical vs. atypical-atypical: $t = 0.08$, $p > .1$, $\beta = 3.06\text{E-}03$, $SE = 3.96\text{E-}02$). The effects of initial-phoneme plosivity, previous RTs and trial number were not significant (initial-phoneme plosivity: $t = 1.88$, $p = .06$, $\beta = 6.55\text{E-}02$, $SE = 3.48\text{E-}02$; previous RTs: $t = 0.78$, $p > .1$, $\beta = 9.65\text{E-}06$, $SE = 1.25\text{E-}05$; trial number: $t = 1.2$, $p > .1$, $\beta = 3.51\text{E-}04$, $SE = 2.93\text{E-}04$). In sum, initial-phoneme plosivity did not influence the results of Experiment 2 significantly.

A Bayesian one-way repeated-measures ANOVA was conducted for the subject data reported in Table 5, using JASP. Setting the default scale r on fixed-effect size of 0.5, for a main effect of word type, BF_{10} (i.e., BF in favor of the alternative) was 0.36. This result suggests inconclusive with respect to whether there was an effect of word type. When Bayesian paired-sample t tests were conducted with the default scale r on effect size of 0.5, for the atypical-atypical words versus atypical-typical words comparison, BF_{10} was 0.334; for the atypical-atypical words versus typical-atypical words comparison, BF_{10} was 0.99; for the atypical-typical words versus typical-atypical words comparison, BF_{10} was 0.52. These analyses of the data of Experiment 2 were inconclusive with respect to whether there were main effects of first-character typicality and second-character typicality. We pursued this further with Bayes factor analyses that compared Experiment 1 with Experiment 2. For the comparison atypical-atypical versus typical-atypical words, the effect was significantly larger in Experiment 1 than in Experiment 2 (83.9 vs. 25.5), $t(62) = 2.85$, $p = .006$; $BF_{10} = 7.15$. For the comparison atypical-

typical versus typical-atypical words, the effect was significantly larger in Experiment 1 than in Experiment 2 (78.0 vs. 17.2), $t(62) = 2.90$, $p = .005$; $BF_{10} = 7.98$. Thus, while we cannot assert with much confidence that the effects of first-position typicality were *eliminated* in Experiment 2, we can say with confidence that, as our hypothesis predicted, the removal of nonwords did reduce the effects of first-character typicality (i.e., the effects of the sublexical kanji reading route on kanji word reading).

Error rates

Error data were analyzed by logistic regression, with word type as a predictor. There was no significant word type effect (typical-atypical vs. atypical-typical: $z = 0.33$, $p > .1$, $\beta = 0.11$, $SE = 0.33$; typical-atypical vs. atypical-atypical: $z = 0.50$, $p > .1$, $\beta = 0.17$, $SE = 0.34$; atypical-typical vs. atypical-atypical: $z = 0.17$, $p > .1$, $\beta = 0.059$, $SE = 0.34$).

Discussion

As we predicted, the absence of nonwords in Experiment 2 resulted in the reduction or elimination of the position of atypicality effect we saw in Experiment 1. This result is in line with prediction from a strategic control hypothesis (e.g., Coltheart, 1978; Cummine et al., 2011; Monsell et al., 1992; Rastle & Coltheart, 1999; Reynolds & Besner, 2008; Woollams, 2005; Zevin & Balota, 2000). The hypothesis postulates that readers can strategically adjust their relative reliance on the two reading routes as a function of the type of material that they are being asked to read. Following the hypothesis, when there were no nonwords present and stimuli were all inconsistent-atypical words, it was strategically advantageous to play down the use of the sublexical reading procedure, and that could be why in our Experiment 2 subjects did not show evidence of using that procedure (i.e., the position of atypicality effect we observed in Experiment 1).

The time-criterion hypothesis was suggested as a hypothesis challenging the strategic control hypothesis by Lupker and colleagues (e.g., Kinoshita & Lupker, 2002; Lupker, Brown, & Colombo, 1997; Taylor & Lupker, 2001). According to the time-criterion hypothesis, participants establish a time criterion to act as a flexible deadline for when articulation should

start. The placement of this deadline is determined by a number of factors such as the average strength of the stimulus-response (S-R) mappings of the stimuli in a block. When all the stimuli in a block are fairly homogeneous with respect to the average strength of the S-R mappings (i.e., pure blocks), the criterion can be set at a point that seems most appropriate for that type of stimulus. On the other hand, when easy and difficult stimuli (i.e., fast and slow stimuli) are mixed together in the same block (i.e., mixed blocks), the criterion tends to get set at a position that is intermediate to the positions used for the fast and slow stimuli. Therefore, articulatory codes for rapidly processed stimuli are allowed to develop beyond the point where articulation could start, whereas the start of articulation for more slowly processed stimuli is initiated while the articulatory codes are still somewhat incomplete. Consequently, in a pure block of fast stimuli (e.g., words), the reading latency would be faster than a mixed block with slow stimuli (e.g., nonwords); conversely, in a pure block of slow stimuli (e.g., nonwords), the reading latency would be slower than a mixed block with fast stimuli (e.g., words). Moreover Lupker et al. (1997) noted that one additional effect of naming slow stimuli more rapidly and naming fast stimuli more slowly should be reciprocal changes in error rates. Naming slow stimuli more rapidly causes larger error rates. On the other hand, naming fast stimuli more slowly does not lead to any improvement in accuracy because error rates for these stimuli tend to be low in the first place.

If all word types are fast stimuli and nonwords are slow stimuli, the time-criterion hypothesis predicts that responses to all word types should be faster in a pure block (Experiment 2) than a mixed block (Experiment 1), but their accuracy should not differ between Experiment 1 and Experiment 2. At least the responses to typical-atypical words should behave in this way because the fastest stimuli were typical-atypical words. In order to test this prediction, we investigated the interaction between word type and experiment in reading latencies (RTs) and error rates. In RTs, the model with the interaction was significantly better than the model without the interaction ($F = 6.65, p < .01$), indicating that there was a significant interaction between word type and experiment. Both atypical-typical and atypical-atypical words were read significantly faster in Experiment 2 than in Experiment 1 (atypical-typical: $t = 3.49, p < .01$; atypical-atypical: $t = 4.04, p < .01$). This experiment effect was not significant for typical-atypical words ($t = 0.44, p > .1$). In error rates, there was a significant effect of experiment. That is, error rates in Experiment 1 were significantly higher than in Experiment 2 ($z = 3.78, p < .01$, $\beta = 5.91E-01, SE = 1.56E-01$). The main effect of Experiment was not modulated by word type significantly because there was not a change in model fit when the interaction between word type and experiment was excluded from a logistic regression model with the interaction and main effects of these factors ($\chi^2 = 0.57, p > .1$). Thus results in both reading

latencies and error rates were inconsistent with predictions from the time-criterion hypothesis.

The comparison between error rates of Experiment 1 and error rates of Experiment 2 also provides evidence supporting the strategic control hypothesis. The sublexical reading procedure produces incorrect phonology for our stimuli (i.e., inconsistent-atypical words), irrespective of position of atypicality. In contrast the lexical reading procedure produces correct phonology. This leads participants to deemphasize information from the sublexical reading procedure and rely on information from the lexical reading procedure. Reading accuracy improves irrespective of position of atypicality consequently. As the strategic control hypothesis predicts, error rates of Experiment 2 were lower than those of Experiment 1 regardless of position of atypicality.

In contrast to our study, a position of irregularity effect was shown even in a condition where all stimuli were words in Roberts et al.'s (2003) study. The difference between Roberts et al.'s results and our results may have come from a methodological difference. They used a reading list that consisted of equal number of regular words and irregular words. On the other hand, the two-character words we used in Experiments 1 and 2 all contained at least one atypical character—one “irregularity.” There were no typical-typical words.

It is possible that the presence of typical-typical kanji words (words which can be correctly read by the nonlexical kanji reading route) might encourage the use of this route just as the absence of any such words (and the absence of nonwords) discouraged the use of this route in Experiment 2. So in Experiment 3 we added to the 78 words used in Experiment 1 a set of typical-typical words. Will this reintroduce the signature effect of nonlexical kanji word reading—the first-character typicality effect—which was present in Experiment 1 and absent (or much reduced) in Experiment 2? This also permitted a fully crossed design: Character Position \times Character Typicality.

Experiment 3

Method

Participants

Twenty-two undergraduate or postgraduate students at University of Tsukuba participated. All were native speakers of Japanese and had normal or corrected-to-normal vision and no reading impairment.

Stimuli

These were the 78 stimuli used in Experiments 1 and 2, plus 26 new stimuli that were two-character kanji words in which

both characters were typical; some were indeed consistent (hereafter typical words; see [Appendix](#)). That is, both characters in each typical word were converted into a corresponding sound in the most common way. The typical words were chosen from the NTT Psycholinguistic Database (Amano & Kondo, 1999) so that mean auditory familiarity, orthographic familiarity, auditory imageability, orthographic imageability, logarithmic frequency, and neighborhood size in each position were matched with the inconsistent-atypical words. Moreover, mean consistency value of typical words in each character position was matched with mean consistency value of typical readings in the atypical-typical words and the typical-atypical words.

These 104 stimuli constituted a crossed 2×2 design in which the first factor was typicality of character-to-sound correspondences in the first position (first-character typicality), and the second factor was typicality of character-to-sound correspondences in the second position (second-character typicality).

We added 54 two-character kanji typical words as fillers. The fillers were also middle-frequency or low-frequency words with logarithmic frequencies at least one standard deviation below the mean of all two-character kanji words.

Apparatus

Unlike Experiments 1 and 2, stimulus presentation and data recording were controlled by the DMDX software (Forster & Forster, 2003) instead of E-Prime 2.0 because initial morae could not be matched across the conditions. Reading aloud responses were recorded directly to the hard drive of the computer at a sampling rate of 22 kHz via a headset microphone fitted to each participant.

Procedure

Participants were seated in front of a computer screen at a distance of approximately 50 cm. They were required to read the presented stimulus aloud as quickly and accurately as possible. Each trial began with a 500-ms presentation of a fixation cross. The fixation cross was replaced by a stimulus (MS Mincho, font size 18) that was visible for 3,000 ms. The intertrial interval was 1,000 ms. There were breaks every 39 trials.

Results

Reading latencies (RTs)

RT measurements were made manually for each recorded utterance via visual inspection of the speech waveform or sound spectrogram with the CheckVocal speech-signal-processing package (Protopoulos, 2007). This can compensate for

confounding effects of initial morae on voice-trigger measurements of reading latencies. The following analyses were conducted for the inconsistent-atypical words available to statistical analyses in Experiment 1 and 2, that is, 22 stimuli per word type, and 22 typical words whose mean word attributes were matched with each type of inconsistent-atypical words. Reading latencies available to statistical analyses were determined by the same procedure as in Experiments 1 and 2. Table 6 shows the mean and standard deviation of reading latencies and error rates (calculated from subject means). A total of 8.2% reading latencies was excluded. This overall exclusion rate was relatively high because the rate of misreadings was relatively high (3.9%).

Logarithmic reading latencies were analyzed by mixed-effects modeling (Baayen, 2008) as follows. Subjects and items were treated as random-effect factors. First-character typicality and second-character typicality were treated as a fixed-effect factor. Trial number and previous RTs (i.e., RT for the preceding trial) were included as a covariate.

Initially we investigated interactions of fixed factors (first-character typicality and second-character typicality) with covariates (trial number and previous RTs). When the mixed-effects Model 1 including the interactions (see Table 7) were compared with a mixed-effects model removing each interaction of fixed factors with either covariate from Model 1, model fit did not improve significantly (First-Character Typicality × Second-Character Typicality × Previous RTs: $\chi^2 = 1.59$, $p > .1$; First-Character Typicality × Second-Character Typicality × Trial Number: $\chi^2 = 0.0077$, $p > .1$; First-Character Typicality × Previous RTs: $\chi^2 = 0.31$, $p > .1$; Second-Character Typicality × Previous RTs: $\chi^2 = 3.63$, $p = .057$; First-Character Typicality × Trial Number: $\chi^2 = 0.30$, $p > .1$; Second-Character Typicality × Trial Number: $\chi^2 = 0.046$, $p > .1$). This means that the First-Character Typicality × Second-Character Typicality interaction and the main effects of these factors were not influenced by previous RTs and trial number significantly.

When the mixed-effects Model 2 that contained the interaction between first-character typicality and second-character typicality and all main effects (see Table 7) was compared with the mixed-effects Model 3 that did not have the interaction (see Table 7), model fit did not improve significantly ($\chi^2 = 1.12$, $p > .1$). This means that there was no significant interaction between first-character typicality and second-character typicality. Using Model 3 that contained only main effects (see Table 7), the main effect of first-character typicality was significant ($t = 3.02$, $p < .01$, $\beta = 8.85\text{E-}02$, $SE = 2.93\text{E-}02$), but the main effect of second-character typicality was not ($t = 0.51$, $p > .1$, $\beta = 1.48\text{E-}02$, $SE = 2.93\text{E-}02$). For covariates, the effect of previous RTs was significant ($t = 2.93$, $p < .01$, $\beta = 2.88\text{E-}05$, $SE = 9.83\text{E-}06$), but the effect of trial number was not ($t = 0.62$, $p > .1$, $\beta = 7.19\text{E-}05$, $SE = 1.16\text{E-}04$).

Table 6 The means and standard deviations of reading latencies and error rates in Experiment 3

		Second-character typicality		
		Typical	Atypical	Mean
First-character typicality	Typical	856.07 (171.68)	890.48 (201.02)	873.00 (181.06)
	Atypical	974.87 (246.60)	961.23 (231.12)	968.94 (236.05)
	Mean	912.89 (203.59)	926.81 (212.15)	
First-character typicality	Typical	0.83 (9.05)	5.79 (23.35)	3.31 (17.88)
	Atypical	6.61 (24.85)	8.26 (27.53)	7.44 (26.23)
	Mean	3.72 (18.92)	7.02 (25.56)	

Note. Standard deviations are noted in brackets

The factor of initial-phoneme plosivity was added as a fixed factor for further analyses. When the mixed-effects Model 1 (see Table 8) that included interactions of fixed factors with covariates (i.e., previous RTs and trial number) was compared with a mixed-effects model removing an interaction with either covariate from Model 1, model fit did not improve significantly (First-Character Typicality \times Second-Character Typicality \times Initial-Phoneme Plosivity \times Previous RTs: $\chi^2 = 0.59, p > .1$; First-Character Typicality \times Second-Character Typicality \times Initial-Phoneme Plosivity \times Trial Number: $\chi^2 = 1.85, p > .1$; First-Character Typicality \times Second-Character Typicality \times Previous RTs: $\chi^2 = 1.66, p > .1$; First-Character Typicality \times Initial-Phoneme Plosivity \times Previous RTs: $\chi^2 = 0.066, p > .1$; Second-Character Typicality \times Initial-Phoneme Plosivity \times Previous RTs: $\chi^2 = 0.88, p > .1$; First-Character Typicality \times Second-Character Typicality \times Trial Number: $\chi^2 = 0.017, p > .1$; First-Character Typicality \times Initial-Phoneme Plosivity \times Trial Number: $\chi^2 = 0.49, p > .1$; Second-Character Typicality \times Initial-Phoneme Plosivity \times Trial Number: $\chi^2 = 1.66, p > .1$; First-Character Typicality \times Previous RTs: $\chi^2 = 0.33, p > .1$; Second-Character Typicality \times Previous RTs: $\chi^2 = 3.61, p = .057$; Initial-Phoneme Plosivity \times Previous RTs: χ^2

$= 0.001, p > .1$; First-Character Typicality \times Trial Number: $\chi^2 = 0.26, p > .1$; Second-Character Typicality \times Trial Number: $\chi^2 = 0.063, p > .1$; Initial-Phoneme Plosivity \times Trial Number: $\chi^2 = 0.62, p > .1$). These nonsignificant interactions were excluded from the following models. When the mixed-effects Model 2 (see Table 8) that included interactions between fixed factors was compared with a mixed-effects model removing each interaction from Model 2, model fit did not improve significantly (First-Character Typicality \times Second-Character Typicality \times Initial-Phoneme Plosivity: $\chi^2 = 0.24, p > .1$; First-Character Typicality \times Second-Character Typicality: $\chi^2 = 1.16, p > .1$; First-Character Typicality \times Initial-Phoneme Plosivity: $\chi^2 = 0.26, p > .1$; Second-Character Typicality \times Initial-Phoneme Plosivity: $\chi^2 = 0.84, p > .1$). This means that there were no significant interactions between the fixed factors. Using the mixed-effects Model 3 (see Table 8) that contained only main effects, the effect of first-character typicality was significant ($t = 3.03, p < .01, \beta = 8.90E-02, SE = 2.94E-02$) but the effect of the second-character typicality was not ($t = 0.52, p > .1, \beta = 1.52E-02, SE = 2.94E-02$). The effect of initial-phoneme plosivity was not significant ($t = 0.69, p > .1, \beta = 2.11E-02, SE =$

Table 7 The mixed-effect models used for analysis in Experiment 3

	Model 1	Model 2	Model 3
First-Character Typicality \times Second-Character Typicality \times Previous RTs	○		
First-Character Typicality \times Second-Character Typicality \times Trial Number	○		
First-Character Typicality \times Previous RTs	○		
Second-Character Typicality \times Previous RTs	○		
First-Character Typicality \times Trial Number	○		
Second-Character Typicality \times Trial Number	○		
First-Character Typicality \times Second-Character Typicality	○	○	
First-character typicality	○	○	○
Second-character typicality	○	○	○
Previous RTs	○	○	○
Trial number	○	○	○

Note. Each model contained interactions or main effects represented by a circle.

Table 8 The mixed-effect models used for additional analysis in Experiment 3

	Model 1	Model 2	Model 3
First-Character Typicality × Second-Character Typicality × Initial-Phoneme Plosivity × Previous RTs	○		
First-Character Typicality × Second-Character Typicality × Initial-Phoneme Plosivity × Trial Number	○		
First-Character Typicality × Second-Character Typicality × Previous RTs	○		
First-Character Typicality × Initial-Phoneme Plosivity × Previous RTs	○		
Second-Character Typicality × Initial-Phoneme Plosivity × Previous RTs	○		
First-Character Typicality × Second-Character Typicality × Trial Number	○		
First-Character Typicality × Initial-Phoneme Plosivity × Trial Number	○		
Second-Character Typicality × Initial-Phoneme Plosivity × Trial Number	○		
First-Character Typicality × Second-Character Typicality × Initial-Phoneme Plosivity	○	○	
First-Character Typicality × Previous RTs	○		
Second-Character Typicality × Previous RTs	○		
Initial-Phoneme Plosivity × Previous RTs	○		
First-Character Typicality × Trial Number	○		
Second-Character Typicality × Trial Number	○		
Initial-Phoneme Plosivity × Trial Number	○		
First-Character Typicality × Second-Character Typicality	○	○	
First-Character Typicality × Initial-Phoneme Plosivity	○	○	
Second-Character Typicality × Initial-Phoneme Plosivity	○	○	
First-character typicality	○	○	○
Second-character typicality	○	○	○
Previous RTS	○	○	○
Trial number	○	○	○

Note. Each model contained interactions or main effects represented by a circle

3.07E-02). The effect of previous RTs was significant ($t = 2.92$, $p < .01$, $\beta = 2.87E-05$, $SE = 9.83E-06$) but the effect of trial number was not ($t = 0.61$, $p > .1$, $\beta = 7.10E-05$, $SE = 1.16E-04$). In sum, the results indicated that initial-phoneme plosivity did not influence the results of Experiment 3 significantly.

A Bayesian two-way repeated-measures ANOVA was conducted for the subject data reported in Table 6, using JASP. Table 9 shows BFs with the default scale r on fixed-effect size of 0.5.

Compared to the null model, BF_{10} for all other models except Model 2 (i.e., the only second-character typicality effect model) exceeds 3.0. The model that receives the most support against the null model was Model 1 (i.e., the only first-character typicality effect model). Adding the interaction to Model 3 (i.e., the two main effects model) decrease the degree of the support for H_1 by a factor $159439.02/119231.16 = 1.34$. This is the Bayes factor in favor of Model 2 versus Model 3. Adding the second-typicality effect to Model 1 (i.e., the only first-character typicality effect model) decrease the degree of the support for H_1 by a factor $513311.06/119231.16 = 4.31$, this result representing the Bayes factor in favor of Model 1 versus Model 3. To confirm these results, Bayesian paired-sample t tests were conducted. Setting the default scale r on effect size of 0.5, for the first-character typicality effect, BF_{10} was 1699;

for the second-character typicality effect, BF_{10} was 0.57; for the interaction, BF_{10} was 0.77. Thus, in Experiment 3, the first-character typicality effect that was clearly present in Experiment 1 but not in Experiment 2 did return, as we had anticipated. The Bayes factor results did not allow us to reach any firm conclusion about whether there was a second-character typicality effect, or about whether there was an interaction between the two factors.

Error rates

Error data were analyzed by logistic regression, with first-character typicality and second-character typicality as predictors. There was neither a significant main effect of first-character typicality ($z = 1.50$, $p > .1$, $\beta = 0.38$, $SE = 0.26$) nor a significant main effect of second-character typicality ($z = 0.98$, $p > .1$, $\beta = 0.24$, $SE = 0.25$). There was a significant interaction between first-character typicality and second-character typicality because the model with the interaction was significantly better than the model without the interaction ($\chi^2 = 11.17$, $p < .01$). The interaction reflected that a main effect of first-character typicality was significant when the second character was pronounced as a typical reading ($z = 4.00$, $p < .01$, $\beta = 2.14$, $SE = 0.53$) but not an atypical reading ($z = 1.50$, $p > .1$, $\beta = 0.38$, $SE = 0.26$), and that a

Table 9 BFs of model comparisons in Experiment 3

Models		BF ₁₀
1	First-typicality	513311.06
2	Second-typicality	0.25
3	First-Typicality + Second-Typicality	159439.02
4	First-Typicality + Second-Typicality + First-Typicality × Second-Typicality	119231.16

main effect of second-character typicality was significant when the first character was pronounced as a typical reading ($z = 3.71$, $p < .01$, $\beta = 2.00$, $SE = 0.54$) but not an atypical reading ($z = 0.98$, $p > .1$, $\beta = 0.24$, $SE = 0.25$).

General discussion

This study revealed that position of an inconsistent-atypical correspondence influenced reading fluency significantly when a reading list contained stimuli for which sublexical processing is not harmful. When participants read inconsistent-atypical words aloud mixed randomly with nonwords, reading latencies of words with an inconsistent-atypical correspondence in the initial position were significantly longer than words with an inconsistent-atypical correspondence in the second position (Experiment 1). The significant position of atypicality effect disappeared when the inconsistent-atypical words were presented alone (Experiment 2). When participants read equal number of inconsistent-atypical words and typical words aloud, the typicality effect was significant in the first-character position but not in the second-character position (Experiment 3). In all experiments, error rates were not influenced by the position of an inconsistent-atypical correspondence.

The position of atypicality effect on reading latencies in this study is analogous to the position of irregularity effect in English. The dual-route approach (e.g., Coltheart et al., 2001) considers the position of irregularity effect in English as evidence that a sublexical reading procedure is applied to words serially. There have been two main challenges to this interpretation. However, both have been examined empirically and have been rejected in previous studies. Our study was also contrary to both challenges as follows.

The first challenge was offered by Kawamoto et al. (1998). They argued that if a stimulus list included words beginning with plosives and nonplosives, a position of irregularity effect can arise because a regularity effect in the initial position is obtained irrespective of initial plosivity, whereas a regularity effect in the second position interacts with initial plosivity (i.e., a significant regularity effect is observed only for words beginning with plosives). But then Cortese (1998) obtained a significant position of irregularity effect irrespective of initial plosivity, suggesting a position of irregularity effect cannot be explained by the different articulation of initial plosivity. In

this study, there was neither a significant main effect of initial plosivity nor a significant interaction between initial plosivity and a fixed factor in all experiments. We believe that initial-phoneme plosivity cannot significantly influence reading latencies in kanji word naming for two reasons. The first is because the initial-phoneme criterion hypothesis of Kawamoto et al. is relevant to the reading of English, whose functional phonological unit is the phoneme, but not relevant to the reading of Japanese, whose functional phonological unit is the mora (Kureta, Fushimi, & Tatsumi, 2006). The second is because almost all our kanji characters corresponded to a sound with more than one mora (i.e., more than two phonemes) irrespective of typicality of character-to-sound correspondence (e.g., for the kanji character 間, the typical reading is /kaN/ and the atypical reading /aida/). Hence, atypical readings in the second-character position did not lie in the second phoneme but rather in later phoneme positions. In other words, in contrast to reading English, in this study, position means the character position rather than the phoneme position.

The second challenge was offered by Plaut, McClelland, Seidenberg, and Patterson (1996) and by Zorzi (2000), the argument being that first-position irregular words tended to be more inconsistent than later-position irregular words in previous studies. But then Roberts et al. (2003) showed a significant position of irregularity effect using stimuli with no confounding between the degree of consistency and serial position. In this study, two-character kanji words were chosen so that the degree of character-to-sound consistency was controlled for both typical and atypical readings separately across character positions. It suggests that there was no confound with character-sound consistency.

Our results might be interpreted in terms of speech production. Meyer (1991) argued that the sublexical units of a word are selected in a particular order, namely, according to their sequence in the word in speech production. This argument is led by her experiments using a speech production task. In the task, participants had to utter one out of three or five response words as quickly as possible in each test trial. For monosyllabic response words, naming latencies were shorter when all words had the same onset (i.e., the homogeneous condition) than when they shared neither the onset nor the rime (i.e., the heterogeneous condition). Such a facilitatory effect was not observed when all response words had the same rime. Similarly, for disyllabic response words, naming latencies

were facilitated more in the homogeneous condition than in the heterogeneous condition when all response words had the same onset of the first syllable but not when they had the same rime of the first syllable. Following these results, it is concluded that syllables are phonologically encoded in two ordered steps, the first of which is dedicated to the onset and the second to the rime. Meyer's study also suggests that naming a word can begin before the entire phonology of the word is not computed, when the initial sound is determined.

One might think that Meyer's (1991) assumption accounts for our results of reading latencies. In general, it takes a longer time to determine the phonology for a character that is pronounced as its atypical reading than it does for a character that is pronounced as its typical reading. If articulation begins as soon as the phonology of the first character is determined, naming words with a typical correspondence in the first position will begin more rapidly than words with an inconsistent-atypical correspondence in the first position, irrespective of whether the sublexical reading process occurs in parallel. This was observed in an analysis of reading latencies of Experiment 1 and 3. Rastle, Harrington, Coltheart, and Palethorpe (2000), however, refuted the idea that the reading-aloud response begins as soon as the initial phoneme is computed. They argued that reading aloud begins when the computation of phonology is complete. Our study used a reading-aloud task but not a speech-production task. In addition, participants had difficulty in anticipating the initial sound of a stimulus before the presentation of the stimuli because the presentation order of our stimuli was randomized, and the initial sounds of our stimuli were various. In sum, our study differed from Meyer's study in terms of the task used and the ease of anticipation of initial sounds. Moreover, the phonological structure of Japanese is completely different from alphabetical languages; nevertheless, there is no previous study investigating the question of whether reading Japanese words aloud begins before the computation of phonology is complete. It is unclear if Meyer's findings can be applied to Japanese.

We obtained a significant position of atypicality effect on reading latencies but not error rates. This is not consistent with previous studies in English, which showed a significant position of irregularity effect on both reading latencies and error rates (e.g., Cortese, 1998; Rastle & Coltheart, 1999; Roberts et al., 2003). These inconsistent results do not conflict with the idea that a sublexical reading procedure is applied to kanji words serially. Our study addressed the reading process of kanji words, while the above previous studies addressed the reading process of English words. English and Japanese kanji are opaque in terms of orthography-to-phonology mapping. However, Japanese kanji is a logographic script, and each kanji character is a morpheme. The characteristics can allow participants to access lexical information easily, relative to English, and thereby use lexical information effectively when they determine phonology of inconsistent-atypical kanji characters. If it is true, then reading accuracy will not be influenced by position of

atypicality, irrespective of whether sublexical reading processing occurs serially. If the sublexical reading procedure is applied to constituents of a kanji word serially, reading latencies will be influenced by the position of atypicality in the same mechanism of the position of irregularity effect in English. Thus, we obtained a significant position of atypicality effect only in analysis of reading latencies.

A reading model for two-character kanji words has been developed within the framework of the PDP models (Plaut et al., 1996; Seidenberg & McClelland, 1989) by Ijuin et al. (1999). Ijuin et al.'s Japanese PDP model computes the phonology on the basis of whole-word patterns and statistical consistencies between kanji characters and their sounds. The model succeeded in simulating the consistency effect observed in Fushimi et al.'s (1999) empirical study. However, whether the model can simulate a position of atypicality effect in our study remains an open question due to the absence of a serially operating mechanism for computing phonology from orthography. We expect that the Japanese PDP model will not succeed in doing this, given that the English PDP models failed to simulate a position of irregularity effect (Roberts et al., 2003).

Alternatively a multilevel interactive activation model of word recognition for two-character kanji words was proposed by Miwa et al. (2014). The model was developed based on empirical data using a lexical decision task. It consists of nodes for visual features, radicals, kanji characters, kanji words, and word meanings. Their model assumes that two-character kanji words are preferentially processed from the initial character to the second character. Miwa et al. did not mention whether the sequential manner lies in the character-to-sound conversion or in visual analyses clearly. If their model worked as the CDP+ model (Perry, Ziegler, & Zorzi, 2007) does (i.e., the serial activation of characters results in phonological assembly in a serial manner), our position of atypicality effect would be explained.

In conclusion, our study shed light on the questions of whether a sublexical reading procedure is applied to the constituent characters of two-character kanji words, and whether this procedure is applied serially or in parallel. Our atypicality effects provide evidence of the existence of a mechanism that translates two-character kanji words from orthography to phonology sublexically (i.e., at the single-character level). Our position of atypicality effects provide evidence that this sublexical procedure accomplishes phonological assembly in a serial rather than parallel manner, a finding which would seem to pose a challenge to approaches assuming that all reading processes work in parallel, such as the PDP models (e.g., Ijuin et al., 1999; Plaut et al., 1996; Seidenberg & McClelland, 1989).

Acknowledgements This work was supported by JSPS KAKENHI Grant Number 15K17420.

Appendix

Table 10 Word stimuli used in Experiment 1, 2, and 3

Inconsistent-atypical words (Experiment 1, 2, and 3)			Typical words (Experiment 3)
<i>Typical-atypical</i>	<i>Atypical-typical</i>	<i>Atypical-atypical</i>	
赤身	空地	海女	礼服
冬鳥	筆先	深酒	代打
冬山	古本	古着	着席
字画	直伝	地声	平然
漢方	下級	書留	急病
寒波	街道	神主	特等
川辺	金具	風下	点線
功名	口語	心得	放流
小言	行楽	声色	統報
黒船	薬屋	首輪	味覚
松林	真顔	真水	習字
南国	長屋	波風	当直
王子	横領	重荷	意表
穩便	遅番	幼心	邪悪
聖者	青果	川柳	文集
先祖	関所	殺生	特産
仕草	質屋	品数	倒立
信者	質素	死神	安易
潮時	下町	白星	通読
退役	台頭	建前	帳面
裏話	雨期	産声	食通
横道	弱虫	夜風	利息
愛想*	間柄*	相性*	番茶*
庭石*	日用*	西日*	和裁*
統率*	遠縁*	冬至*	外界*
開口*	角地*	活甕*	内野*

Note. * indicates the stimuli that were excluded from the statistical analyses

References

- Amano, S., & Kondo, K. (1999). *The NTT Japanese Psycholinguistic Database*. Tokyo, Japan: Sanseido.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. New York, NY: Cambridge University Press.
- Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp.151–216). London, UK: Academic Press.
- Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human, Perception, and Performance*, *20*, 1197–1211.
- Coltheart, M., Rastle, K., Perry, C., Ziegler, J. & Langdon, R. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–256.
- Cortese, M. J. (1998). Revisiting serial position effects in reading. *Journal of Memory and Language*, *39*, 652–665.
- Cortese, M. J., & Simpson, G. B. (2000). Regularity effects in word naming: What are they? *Memory and Cognition*, *28*, 1269–1276.
- Cummine, J., Amyotte, J., Pancheshen, B., & Chouinard, B. (2011). Evidence for the modulation of sublexical processing in go no-go naming: The elimination of frequency × regularity interaction. *Journal of Psycholinguistic Research*, *40*, 367–378.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods Instruments and Computers*, *35*, 116–124.
- Fushimi, T., Ijuin, M., Patterson, K., & Tatsumi, I. (1999). Consistency, frequency, and lexicality effects in naming Japanese kanji. *Journal of Experimental Psychology: Human, Perception, and Performance*, *25*, 382–407.
- Ijuin, M., Fushimi, T., Patterson, K., & Tatsumi, I. (1999). A connectionist approach to Japanese kanji word naming. *Psychologia*, *42*, 267–280.
- Kawamoto, A. H., Kello, C. T., Jones, R., & Bame, K. (1998). Initial phoneme versus whole-word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*, 862–885.

- Kinoshita, S., & Lupker, S.J. (2002). Effects of filler type in naming: Change in time criterion or attentional control pathways? *Memory & Cognition*, *30*, 1277–1287.
- Kureta, Y., Fushimi, T., & Tatsumi, I. (2006). The functional unit in phonological encoding: Evidence for moraic representation in native Japanese speakers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1102–1119.
- Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 570–590.
- Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, *30*, 69–89.
- Miwa, K., Libben, G., Dijkstra, T., & Baayen, H. (2014). The time-course of lexical activation in Japanese morphographic word recognition: Evidence for a character-driven processing model. *The Quarterly Journal of Experimental Psychology*, *67*, 79–113.
- Monsell, S., Patterson, K. E., Graham, A., Hughes, C. H., & Milroy, R. (1992). Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 452–467.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*, 273–315.
- 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*, 56–115.
- Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*, 859–862.
- Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human, Perception, and Performance*, *25*(2), 482–503.
- Rastle, K., Harrington, J., Coltheart, M., & Palethorpe, S. (2000). Reading aloud begins when the computation of phonology is complete. *Journal of Experimental Psychology: Human, Perception, and Performance*, *26*, 1178–1191.
- Reynolds, M., & Besner, D. (2008). Contextual effects on reading aloud: Evidence for pathway control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 50–64.
- Roberts, M. A., Rastle, K., Coltheart, M., & Besner, D. (2003). When parallel processing in visual word recognition is not enough: New evidence from naming. *Psychonomic Bulletin & Review*, *10*, 405–414.
- Saito, H., Masuda, H., & Kawakami, M. (1998). Form and sound similarity effects in kanji recognition. *Reading and Writing*, *10*, 323–357.
- Seidenberg, M. S., & McClelland, J. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.
- Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior*, *23*, 383–404.
- Taylor, T. E., & Lupker, S. J. (2001). Sequential effects in naming: A time-criterion account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 117–138.
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*. Advance online publication. <https://doi.org/10.3758/s13423-017-1323-7>
- Woollams, A. M. (2005). Imageability and ambiguity effects in speeded naming: Convergence and divergence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 878–890.
- 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*, 1155–1168.
- Zevin, J. D., & Balota, D. A. (2000). Priming and attentional control of lexical and sublexical pathways during naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 121–135.
- Zorzi, M. (2000). Serial processing in reading aloud: No challenge for a parallel model. *Journal of Experimental Psychology: Human, Perception, and Performance*, *26*, 847–856.