



Phonological priming effects with same-script primes and targets in the masked priming same-different task

Huilan Yang¹ · Masahiro Yoshihara² · Mariko Nakayama³ · Giacomo Spinelli⁴ · Stephen J. Lupker⁴

Published online: 24 August 2020
© The Psychonomic Society, Inc. 2020

Abstract

Norris, Kinoshita and colleagues (Kinoshita & Norris, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1), 1–18, 2009; *Journal of Experimental Psychology: General*, 137(3), 434–455, 2010; Norris & Kinoshita, *Quarterly Journal of Experimental Psychology*, 63(1), 194–204, 2008) have suggested that the masked priming same-different task (SDT) is an excellent tool for studying the orthographic coding process because, in most circumstances, performance in that task is driven entirely by orthographic codes. More specifically, although evidence of phonological influences (i.e., phonological priming effects in the SDT) have been reported, Kinoshita, Gayed, and Norris (*Journal of Experimental Psychology: Human Perception and Performance*, 44(11), 1661–1671, 2018) have claimed that phonological priming does not arise in the SDT when the prime and target are written in the same script and the targets are words, the most typical experimental situation. Indeed, it does appear that no-one has yet reported phonological priming effects in such situations. The question of whether it is possible to observe phonological priming in such situations was more fully examined in the present experiments. Experiment 1 involved a masked priming SDT using Japanese Kanji script in which the primes and targets were homophonic but shared no characters. Experiment 2 was a parallel experiment using Chinese stimuli. In both experiments, phonological priming effects were observed for both one- and two-character words. These experiments indicate that, although the priming effects in masked priming SDTs undoubtedly have a strong orthographic basis, phonological codes also play a role even when the prime and (word) target are written in the same script.

Keywords Masked priming same-different task · Phonological · Homophonic · Orthographic · Japanese · Chinese

✉ Stephen J. Lupker
lupker@uwo.ca

Huilan Yang
yanghuilanhn@hotmail.com

Masahiro Yoshihara
m.yoshihara@aoni.waseda.jp

Mariko Nakayama
mariko.nakayama.d5@tohoku.ac.jp

Giacomo Spinelli
gspinel@uwo.ca

¹ Department of Foreign Languages, Zhejiang Gongshang University, Hangzhou 310018, Zhejiang, China

² Faculty of Letters, Arts, and Sciences, Waseda University, 1-24-1 Toyama, Shinjuku-ku, Tokyo, Japan

³ Division of Language Research, Department of International Cultural Studies, Tokohu University, 41 Kawauchi Aoba-ku, Sendai, Miyagi, Japan

⁴ Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2, Canada

Introduction

Orthographic coding refers to the component of the reading process that produces a representation reflecting both the letter identities and their positions in the word being read. Successful completion of this process is quite important in reading as, otherwise, readers could not distinguish orthographically similar words like “trial” and “trail.” The experimental paradigm most commonly used in investigations of this process is the masked priming paradigm. In this paradigm, a prime is presented for a brief period (e.g., 50 ms), so that, in general, participants cannot identify the prime or even notice its existence, followed by a target to which participants must make a response. The most typical response is a lexical decision (i.e., word-nonword) response (Forster & Davis, 1984).

In investigations of orthographic coding, the prime and target will have some orthographic relationship between them (e.g., honse-HOUSE) and the size of the priming effect is typically taken as a measure of the degree of orthographic similarity of the prime and target. Researchers have assumed that by varying the nature of the orthographic relationship

between the two stimuli and noting the size of the priming effect that is produced, the nature of orthographic coding will become better understood. Indeed, a number of relevant findings have emerged. For instance, developing readers produced significant honse-HOUSE type priming effects for words with large orthographic neighborhoods (Coltheart, Davelaar, Jonasson, & Besner, 1977), whereas skilled adult readers did not (Castles, Davis, & Letcher, 1999). As a second example, transposed-letter nonwords (hosue-HOUSE) produce larger priming effects than substituted-letter nonwords (honae-HOUSE) in many different languages (Perea & Lupker, 2003, 2004; Yang, Chen, Spinelli, & Lupker, 2019).

There are, however, some limitations to the use of this basic technique. One is that the masked priming LDT has also been shown to be affected by phonological (Ferrand & Grainger, 1992, 1993) and lexical (Davis & Lupker, 2006) information. In an attempt to provide a way of examining orthographic coding independent of phonological and lexical (and other) factors, Norris and Kinoshita (2008) introduced the masked priming same-different task (SDT). In this task, participants will see a reference stimulus above a forward mask (e.g., #####) for 1,000 ms followed by a prime for 50 ms in the same position as the mask had been and then a target also in that same position. The participants' task is to decide whether the target is the same as or different from the reference stimulus. Just like in the masked priming LDT, the priming effects in the masked priming SDT seem to be invariant with respect to changes in visual inputs (e.g., font, size, and uppercase/lowercase; García-Orza, Perea, & Muñoz, 2010; García-Orza, Perea, & Estudillo, 2011; Kinoshita & Norris, 2009). More importantly, the priming effects in this task have also been found to be independent of target frequency, lexicality, and morphology (Duñabeitia, Kinoshita, Carreiras, & Norris, 2011; Kinoshita & Norris, 2009), suggesting that effects in the masked priming SDT might be purely orthographic (Kinoshita & Norris, 2009, 2010; Norris & Kinoshita, 2008).

If this conclusion is correct, it would have very important implications for the investigation of most current theories of orthographic coding (e.g., Davis, 2010; Gómez, Perea, & Ratcliff, 2008; Grainger, Granier, Farioli, Van Assche, & Van Heuven, 2006; Grainger & Van Heuven, 2003; Norris & Kinoshita, 2012; Norris, Kinoshita, & Van Caasteren, 2010; Schoonbaert & Grainger, 2004; Whitney, 2001; Whitney & Marton, 2013). Specifically, according to virtually all of these theories, the degree to which any given lexical representation is activated by reading a word is a function of the similarity of the orthographic code that is produced by reading the word to the orthographic information stored in the word's lexical representation. Similarly, in a masked priming situation, the assumption is that the orthographic code produced by the masked prime will activate the lexical representation of a word to the extent that the orthographic information contained in that word's lexical representation is similar to the orthographic code produced by prime processing. Therefore, if it can legitimately be assumed that no other factors

affect the priming process in a particular task, the size of the observed priming effect would document the similarity of the prime's and the target's orthographic codes. As such, if it can be demonstrated that, at least in certain situations, performance in the SDT is totally an orthographic phenomenon, it would clearly make that task the optimal tool for contrasting the various theories of orthographic coding, in particular, the optimal tool for contrasting the different assumptions that those theories make about the nature of the orthographic code.

Although it is now fairly clear that the SDT is unaffected by many nonorthographic factors, the question of whether there are phonological influences in the SDT has been somewhat more difficult to resolve for the reason that, in many languages, particularly in alphabetic languages, it is difficult to tease apart the effects of orthography and orthographically driven phonology. That is, in those languages, most letters/characters are associated with only a single speech sound and many speech sounds can only be represented by a single letter/character or, in alphabetic languages, bigram. Hence, any priming effects in alphabetic language experiments ostensibly produced by orthographic similarity could have been due either to phonological similarity or, more likely, to some interaction of orthographic and phonological influences. If either of these possibilities were to be true, the implication would be that the priming effects observed in the masked priming SDT are providing somewhat less than an uncontaminated view of the orthographic coding process.

In one attempt to address the question of a potential influence of phonology in the masked priming SDT, Kinoshita and Norris (2009), using native English speakers, reported that repetition primes (e.g., score) facilitated target (e.g., SCORE) processing more than pseudohomophone (e.g., skore) or one-letter-different (ILD) primes (e.g., smore), and there was no significant difference between the latter two prime types. The lack of a difference between skore- and smore-type primes was taken by those authors as suggesting that phonology does not play a role in the SDT.

The difficulty with drawing such a conclusion from this result, however, is that, as Lupker, Nakayama, and Perea (2015) point out, Kinoshita and Norris's (2009) manipulation of phonology is a weak one. That is, the phonological distinction between the primes skore and smore is quite small (i.e., a single phoneme, with skore matching SCORE at four phoneme positions and smore matching at three phoneme positions). Further, both primes match SCORE orthographically at four letter positions. Therefore, both skore and smore should be able to provide considerable priming at both the orthographic and the phonological levels. In such a situation, there is a reasonable possibility that the single phoneme difference between skore and smore would not alter the priming effect to any measureable degree.

Certainly, as Kinoshita et al. (2018) have noted, there is at least one situation in the literature in which a single phoneme difference can produce an effect in a masked priming

paradigm. That is, a difference in the prime's onset phoneme (either matching or mismatching that of the target) is enough to produce significantly different target latencies in a naming task (e.g., Forster & Davis, 1991; Kinoshita, 2000). What's important to realize, however, is that performance in a task of that sort is explicitly based on phonological information with a special emphasis on the target's onset phoneme (the sound that triggers the voice key), a situation quite unlike the situation in the SDT. The bottom line, therefore, is that if one is going to be able to determine whether there is also a phonological component in the SDT, one needs to create a somewhat stronger manipulation of phonology. Because the problem of creating a reasonably strong manipulation of phonology independent of the effects of orthography is difficult, if not impossible, to overcome in alphabetic languages, Lupker and colleagues (Lupker et al., 2015; Lupker, Nakayama, & Yoshihara, 2018) took a different approach, examining *cross-script* priming effects in the SDT and, in doing so, demonstrated clear phonological priming effects.

More specifically, Lupker et al. (2015) showed that Japanese-English bilinguals produced priming effects in the SDT using English reference stimuli and targets with Japanese Katakana primes (e.g., reference stimulus: south, prime: サオス and, target: SOUTH, where サオス is a nonword in Japanese that is phonologically similar to SOUTH). As there is no orthographic overlap between Japanese Katakana and English, the priming effect observed by Lupker et al. (2015) is most likely phonologically based. In a follow-up, Lupker et al. (2018) were able to show a similar effect using different script, but within language, primes and targets. That is, Lupker et al. (2018) showed priming effects using Kanji reference stimuli and targets with Hiragana transcription primes (e.g., reference stimulus: 記号, prime: きごう, target: 記号). These results provide additional support for the claim that even though the priming in the masked priming SDT has a considerable orthographic component, phonology does play a role even when the stimuli are familiar words from a reader's first language.¹

In their more recent follow-up, however, Kinoshita et al. (2018) took issue with this claim.² Kinoshita et al. used the same manipulation as Kinoshita and Norris (2009) with a new set of

stimuli created by changing the first letter in a target word in Experiment 1 and in a target nonword in Experiment 2. In Experiment 1, repetition primes (e.g., cult) again facilitated target processing to a larger degree than pseudohomophone (e.g., kult) or 1LD primes (e.g., nult), which showed, again, a 2-ms difference (i.e., this particular manipulation again produced little evidence of phonological priming in English for word targets in their Experiment 1). There was, however, evidence of phonological priming in their Experiment 2 (for nonword targets) in which the pseudohomophone primes produced a latency that was 13 ms shorter than that for the 1LD primes. In the end, Kinoshita et al. concluded that: (1) phonological priming effects when primes and targets are from the same writing system are different than when they come from different writing systems and, in particular, (2) that phonology does not play a role in producing priming effects in the masked priming SDT when the prime and target are written in the same script, at least when words are being used as targets (as in their Experiment 1). That is, priming effects with same script primes and (word) targets are, instead, purely orthographically based.

Kinoshita et al. (2018) also suggested that the reason that evidence of phonological priming had been observed when the primes and targets are in different scripts (i.e., when the primes do not share orthography with the targets, e.g., Lupker et al., 2015, 2018) is because there is no orthographic competition between the prime and target letter identities in that situation, allowing phonology to have an impact. They then extended this argument to explain the phonological priming effect for (same-script) nonword targets in their Experiment 2. That is, they suggested that their phonological priming effect for nonwords may have the same basis as the phonological priming effect observed with primes and targets written in different writing systems (i.e., the competition between prime and target letter identities is minimal when the targets are nonwords).

Kinoshita et al.'s (2018) claim that phonological priming has not been observed in the masked priming SDT when the primes and (word) targets are written in the same script does appear to be consistent with the extant data. Further, the claim is a relevant one because most of the masked priming SDT experiments now in the literature, experiments which have been used to draw conclusions about orthographic processing, have been done using same-script primes and targets. Further, most of those experiments have involved word targets, although, as noted previously, the priming effects observed in those previous experiments did not appear to be affected by the lexical status of the targets, unlike the pattern Kinoshita et al. (2018) observed.

What is also true, however, is that, as noted previously, to this point, no-one has looked at phonological priming in the masked priming SDT involving same-script primes and targets using a strong manipulation of phonology. Because, as discussed, it is virtually impossible in alphabetic languages to

¹ Kinoshita, Schubert, and Verdonschot (2019) also appear to have demonstrated cross-script phonological priming in an SDT (in Japanese), using Kana (both Katakana and Hiragana) reference stimuli and targets along with homophonic Kanji primes. However, there is no report in that paper as to whether the 17-ms priming effect was significant.

² Actually, it is not entirely clear what "claim" Kinoshita et al. (2018) took issue with. In their paper, they state, "Lupker et al. (2015) reported a finding using the same-different task... which, they suggested, implies that previous reported priming effects with the Roman alphabet may have been due to phonological, rather than, orthographic, similarity" (p. 1662). That quote, however, does not accurately reflect Lupker et al.'s position. Their position was that such effects likely resulted from a combination of the impact of phonological *and* orthographic similarity. As described just below, however, Kinoshita et al. do appear to have taken issue with that position as well, at least in situations when the targets are words.

create a strong manipulation of phonology that is independent of orthography, in the present experiments this issue was investigated using logographic scripts, specifically, Japanese Kanji and Chinese. In these scripts, there are many homophonic characters (e.g., 紅 (/hóng/, red) is a homophone of 宏 (/hóng/, big)) and those characters are, otherwise, completely different from one another (i.e., both orthographically and semantically). Therefore, it is possible to create homophonic character strings that share no characters (or meaning) in both Japanese Kanji and Chinese. Both of these scripts were used in the present experiments. Using such a strong manipulation of phonological relatedness (i.e., using primes and targets that completely match in phonology) should create the optimal test of whether masked priming effects in the SDT truly reflect only the impact of the orthographic coding process or those effects also reflect the impact of phonological information.

Experiment 1 involved both one- and two-character Kanji words as reference stimuli, primes, and targets (with Japanese native speakers) in a masked priming SDT. Experiment 2 involved one- and two-character Chinese words as reference stimuli, primes, and targets (with Chinese native speakers as participants), using a procedure that paralleled the procedure used in Experiment 1. If priming effects in this task are at all phonological, the homophone conditions should facilitate target responses on the “same” trials relative to those in the unrelated condition. However, if priming in this task is purely orthographic due to the fact that the primes and (word) targets are written in the same script, there should be no homophone priming effect because neither the homophone primes nor the unrelated primes are orthographically similar to their targets.

Word length was included as a factor essentially because most of the experiments examining phonological processing in Chinese in the past have tended to do so using one-character words (Perfetti & Tan, 1998; Perfetti & Zhang, 1995; Zhou & Marslen-Wilson, 1999), whereas most of the experiments involving phonological processing in Japanese have tended to do so using two-character words (e.g., Fushimi, Ijuin, Patterson, & Tatsumi, 1999; Hino, Kusunose, Lupker, & Jared, 2013; Hino, Miyamura, & Lupker, 2011; Tamaoka, 2007; Wydell, Patterson, & Humphreys, 1993). If phonological priming effects do exist in the SDT, presumably they would not be limited to either one-character or two-character words in either language.

One final point to note is that the general consensus has been that phonological activation is slower when reading in logographic scripts than when reading in alphabetic scripts (e.g., Perfetti, Liu, & Tan, 2005; Perfetti, Zhang, & Berent, 1992). In fact, Li, Rayner, and Cave (2009) have suggested that, when reading in Chinese, phonology is activated so slowly that it plays virtually no role in the reading process in general. Therefore, any phonological effects produced here by our logographic primes would be expected to be somewhat small, certainly smaller than the effects that would be obtained if we had been able to carry out parallel experiments in an alphabetic script language.

Method – Experiment 1

Participants Thirty-six Japanese native speakers from Waseda University (Tokyo, Japan) participated in this experiment. They all received 500 yen (about US\$4) for their participation and indicated that they had normal or corrected-to-normal vision with no known reading disorder.³

Materials One hundred and twenty Kanji stimuli were chosen as targets for the “same” trials. Half of the targets were one-character Kanji words and the other half were two-character Kanji words. For the one-character Kanji targets, the mean character frequency (per 570,554,885) was 34,131 (range: 349–315,932) and for the two-character Kanji targets, the mean word frequency (per 287,792,797) was 2,092 (range: 16–17,115) according to Amano and Kondo (2003). Although a single Kanji often has multiple pronunciations, care was taken to make sure that all of the single Kanji character stimuli used in this experiment had only one pronunciation (according to Amano & Kondo, 2003).

We selected two types of primes for each target: (1) a homophonic prime; (2) an unrelated prime. Homophonic primes (e.g., reference stimulus: 副/fuku/, prime: 服/fuku/, target: 副/fuku/, reference stimulus: 改名/kaimei/, prime: 解明/kaimei/, target: 改名/kaimei/) were primes that had the same phonology (and character length) as their targets/reference stimuli, with there being no character or semantic overlap between the primes and targets. The unrelated primes had no phonological, character, or semantic overlap with their targets (e.g., reference stimulus: 副/fuku/, prime: 審/shiN/, target: 副/fuku/, reference stimulus: 改名/kaimei/, prime: 税率/zeiritsu/, target: 改名/kaimei/). The word targets were divided into two counterbalanced lists with each list containing 30 stimuli in each condition. Half of the participants were assigned to one list, and the other half were assigned to the other list.

For the “different” trials, another set of 120 Kanji stimuli (60 one-character Kanji and 60 two-character Kanji words) was also selected as targets. For the one-character Kanji targets, the mean character frequency (per 570,554,885) was 31,281 (range: 1,046–340,688), and for the two-character Kanji targets, the mean word frequency (per 287,792,797) was 2,110 (range: 51–15,141) according to Amano and Kondo (2003). In addition, a different set

³ Our selection of the sample sizes for the two experiments was based on two considerations. First, we felt it necessary to use sample sizes (36 in Experiment 1, 54 in Experiment 2) that were at least as large as those used by Kinoshita et al. (2018) (30 in their Experiment 1, 31 in their Experiment 2). In addition, in order to address any more specific issues concerning power, we felt it necessary to make sure that the relevant means were based on at least 1,600 data points (Brybaert & Stevens, 2018). Because some trials would undoubtedly be lost as a result of errors and outliers, we compensated for that fact by having 2,160 trials (36 × 60) trials per mean in Experiment 1 and 3,240 trials (54 × 60) per mean in Experiment 2. Note also that power analyses were done for both experiments with their results reported in Tables 2 and 3 for Experiment 1 and Tables 5 and 6 for Experiment 2.

of 120 Kanji stimuli were selected to serve as reference stimuli: 60 of them were one-character Kanji words (character frequency, $M = 34,048$ per 570,554,885, range = 971–562,593) and the other 60 were two-character Kanji words (word frequency, $M = 2,108$ per 287,792,797, range = 62–15,120). The reference stimuli were orthographically, phonologically, and semantically unrelated to their targets (thus yielding “different” responses).

The homophonic and unrelated primes were set up in a similar way as was done for the “same” trials; however, only one list of stimuli was used. For the one-character Kanji targets, 30 of them were preceded by a homophonic prime (e.g., reference stimulus: 症/shou/, prime: 每/mai/, target: 枚/mai/) and the other 30 by an unrelated prime (e.g., reference stimulus: 艇/tei/, prime: 特/toku/, target: 墨/ruì). Similarly, half of the two-character Kanji targets were preceded by a homophonic prime (e.g., reference stimulus: 挑凳/chouhatsu/, prime: 家庭/katei/, target: 假定/katei/), and the other half by an unrelated prime (e.g., reference stimulus: 色素/shikiso/, prime: 齋場/saijou/, target: 謙虛/kenkyo/). The stimuli for both experiments can be found in Appendix 1.

Procedure DMDX (Forster & Forster, 2003) software was used to control stimulus presentation and data collection. The stimuli were presented on a 15-in. CRT monitor using a refresh rate of 60 HZ (16.67 ms). The screen resolution was 1,024 × 768. The experimental materials were all presented in 12-pt Arial Unicode font.

The sequence of each trial was: a row of hashtags (#####) presented below a reference stimulus for 1,000 ms, followed by a prime for 50 ms in the same position as the row of hashtags and then the target in that same position as the prime for 3,000 ms or until the participant responded. Participants were asked to decide whether each target was the same as or different from the reference stimulus and press the “SAME” button on a response box if they are the same and the “DIFFERENT” button on a response box if they are different. They were asked to respond as quickly and as accurately as possible. Stimulus presentation was randomized for each subject. The experimental block included 240 trials in total, 120 “same” trials and 120 “different” trials. Participants received 16 practice trials before starting the experimental block. This experiment was approved by the Waseda University Research Ethics Board (Protocol # 2018-216).

Method - Experiment 2

Participants Thirty-eight native Chinese speakers from Western University (London, Ontario, Canada) and another 16 native Chinese speakers from Zhejiang Gongshang University (Hangzhou, Zhejiang, China) participated in this experiment. Participants from Western University received course credit for their participation, and participants from

Zhejiang Gongshang University received 5 Chinese dollars for their participation. They all indicated that they were highly proficient in reading simplified Chinese and had normal or corrected-to-normal vision with no reading disorder.

Materials One hundred and twenty simplified Chinese words (60 one-character Chinese words and another 60 two-character Chinese words) were chosen as the reference stimulus/target words on the “same” trials.⁴ The mean word frequency (per million) of these one-character Chinese words is 442.61 (range: 0.83–11,853.78), and the mean word frequency (per million) of these two-character Chinese words is 60.62 (range: 0.06–1,824.99) according to the *SUBTLEX-CH* database (Cai & Brysbaert, 2010).

As in Experiment 1, we selected two types of primes for each word target: (1) a homophone prime and (2) an unrelated prime. Homophone primes (e.g., reference stimulus: 红/hóng/, prime: 宏/hóng/, target: 红/hóng/; reference stimulus: 歧视/qí shì/, prime: 骑士/qí shì/, target: 歧视/qí shì/) are primes that have the same phonology as the targets, with there being no character or semantic overlap between the two character strings. Unrelated primes had no phonological, character, or semantic overlap with their targets (e.g., reference stimulus: 红/hóng/, prime: 到/dào/, target: 红/hóng/; reference stimulus: 歧视/qí shì/, prime: 香蕉/xiāng jiāo/, target: 歧视/qí shì/). The counterbalancing was identical to that in Experiment 1.

In addition, a different set of 240 Chinese words was selected to serve as reference stimuli and targets on “different” trials. Half were one-character Chinese words and the other half were two-character Chinese words. The mean word frequency (per million) of these one-character Chinese target words in the *SUBTLEX-CH* database (Cai & Brysbaert, 2010) is 182.85 (range: 0.12–2,034.22), and the mean word frequency (per million) of these two-character Chinese target words is 143.75 (range: 0.15–915.24). The mean word frequency (per million) of these one-character Chinese reference words in the *SUBTLEX-CH* database (Cai & Brysbaert, 2010) is 113.55 (range: 96.49–128), and the mean word frequency (per million) of these two-character Chinese reference words is 109.69 (range: 95.69–132.68). The reference stimuli were orthographically, phonologically, and semantically unrelated to their targets; however, they were the same length as their targets. The homophone and unrelated primes were set up in a similar way as those for the “same” trials (e.g., homophone primes – reference stimulus: 集/jí/, prime: 保/bǎo/, target: 饱/bǎo/; reference stimulus: 打扰/dǎ rǎo/, prime: 冒进/mào jìn/, target: 毛巾/máo jīn/; unrelated primes – reference stimulus: 满/mǎn /, prime: 声/shēng/, target:

⁴ Most of the stimuli used in Experiment 2 were drawn from the stimuli used by Perfetti and Tan (1998) and by Zhou and Marslen-Wilson (2000).

间/jiàn/; reference stimulus: 姑娘/gū niang/, prime: 目标/mù biāo/, target: 世界/shì jiè/).⁵ However, only one list of stimuli was used for the “different” trials.

Procedure E-prime 2.0 software (Psychology Software Tools, Pittsburgh, PA, USA; see Schneider, Eschman, & Zuccolotto, 2002) was used for data collection. The stimuli were presented on a 19-in. CRT monitor using a refresh rate of 60 Hz (16.67 ms). The screen resolution was 1,280 × 960.

There were three procedural differences between Experiment 2 and Experiment 1, one of which was that responses were made using a keyboard attached to the computer with participants being asked to press the “J” button if the reference stimulus and the target are the same and the “F” button if they are different. Another procedural difference was that participants received eight practice trials before starting the experimental block, instead of 16 practice trials. Third, the primes and targets used different font styles and sizes (35-pt Boldface font for the primes and 40-pt Song font for the targets). The trial sequence was identical to that of Experiment 1. This experiment was approved by the Western University REB (Protocol # 108835).

Results

Correct response latencies less than 250 ms or more than 3 standard deviations from the participant’s mean latency (1.7% of the data in Experiment 1, 1.6% of the data in Experiment 2) were excluded from the latency analyses. The data from “different” trials were not analyzed due to the fact that those trials were not counterbalanced across prime type.

In order to provide as comprehensive an evaluation of the potential priming effect as possible, the latency data in each experiment were analyzed using five different techniques. In all of these techniques both Word length (one-character vs. two-character words) and Relatedness (phonologically related vs. phonologically unrelated) were treated as fixed effects, whereas subjects and/or items were treated as random effects (Baayen, 2008; Baayen, Davidson, & Bates, 2008). In addition, the relevant counterbalancing factor (groups/sets) was also included as a fixed effect to account for variance associated with the participant groups and word sets created for counterbalancing (Pollatsek & Well, 1995). Effects involving that factor are of no importance to the main questions and will

not be reported. The first two techniques were the conventional F_s and F_i techniques, techniques that are regarded as being a bit less sensitive than the mixed-effects models that have now become more popular. The third and fourth techniques were linear mixed-effects model (LMM) techniques. In the third technique, raw reaction times (RTs) were analyzed whereas in the fourth technique, a reciprocal transformation (e.g., $\text{invRT} = -1,000/\text{RT}$) was used in order to normalize the RT distributions. The fifth technique was a generalized linear mixed-effects model (GLMM) technique based on raw RTs. For the error data, only ANOVAs and GLMMs were conducted because LMM analyses require that the data be reasonably well described by a normal distribution and error data are binomial data.⁶

All analyses were conducted in R version 3.5.1 (R Core Team, 2018). Prior to running the model, R-default treatment contrasts were changed to sum-to-zero contrasts (i.e., contr.sum) to help interpret lower-order effects in the presence of higher-order interactions (Levy, 2014; Singmann & Kellen, 2019). ANOVAs were run using the `aov` function in base R. LMMs were run using the `lmerTest` function in the `lmerTest` package, version 3.0-1 (Kuznetsova, Brockhoff, & Christensen, 2017). GLMMs were run using the `glmer` function in the `lme4` package, version 1.1-23 (Bates, Mächler, Bolker, & Walker, 2015). For GLMMs, a Gamma distribution was used to fit the raw RTs, with an identity link between fixed effects and the dependent variable (Lo & Andrews, 2015) and a binomial distribution was used to fit the error data, with a logit link between the fixed effects and the dependent variable.

Estimates of effect size were obtained, for ANOVAs, by calculating η_p^2 for each effect using the `eta_sq` function in the `sjstats` package, version 0.18 (Lüdtke, 2020). For LMMs and GLMMs, we calculated semipartial R^2 for each fixed effect (i.e., the proportion of variance explained by each fixed effect) using the `r2beta` function in the `r2glmm` package, version 0.1.2 (Jaeger, Edwards, Das, & Sen, 2017) with the method proposed by Nakagawa and Schielzeth (2013) and later modified by Johnson (2014). Finally, we conducted power analyses to determine the observed power for each effect. For ANOVAs, power was determined using G*Power software, version 3.1.9.6 (Faul, Erdfelder, Buchner, & Lang, 2009). For LMMs, power was determined using the `powerSim` function in the `simR` package, version 1.0.5 (Green & MacLeod, 2016;

⁵ When designing “different” trials in the SDT, one can choose to create related trials by having the primes related to either the target or the reference stimulus. Perea, Moret-Tatay, and Carreiras (2011) refer to the latter approach as the zero-contingency procedure and have shown that, although this procedure does alter the data pattern on “different” trials, it does not affect the data pattern on “same” trials (see also Kinoshita & Norris, 2010). In both of the present experiments, the related condition on “different” trials was created by selecting a prime that was related to the target rather than to the reference stimulus. This procedure was also the procedure used by Kinoshita et al. (2018).

⁶ Calculating our statistics in five different ways is, of course, not a standard procedure. However, as many of us have discovered, the different procedures can yield different results because the different procedures provide different levels of power for any given sample size. Hence, depending on whether one is trying to get a null effect or a significant effect, one may be able to find an analysis that gives them the answer they’re looking for. Because we wanted to avoid suspicion that something like that happened here, we chose to report the results from all the main analysis techniques that one currently finds in the literature.

see also Brysbaert & Stevens, 2018) in R. The latter series of power analyses was conducted by comparing, for each effect, the full model with the model without that effect (and the interactions that effect was involved in) using a likelihood-ratio test and performing 1,000 simulations for this comparison. Likely due to the complexity of our GLMMs, simulations failed in all cases for those models. Therefore, we report no power analyses for GLMMs in either the latency or error analyses.

In the current version of lme4, convergence failures in the basic analysis involving mixed-effects models, especially GLMMs, are frequent, although many of those failures reflect false positives (Bolker, 2020). To limit the occurrence of convergence failures, we kept the random structure of the mixed-effects models as simple as possible by using only random intercepts for subjects and items. The maximum number of evaluations in model estimation was also increased to 1,000,000 in GLMMs as the default number (i.e., 10,000) is sometimes insufficient for convergence in those models. Even so, GLMMs failed to converge in all cases in the latency analyses. However, convergence was obtained when model estimation was restarted from the apparent optimum (as per the recommended troubleshooting procedure, see convergence help page in R). We report the results from the GLMMs that managed to converge. Convergence warnings were still issued when those models were submitted to semipartial R^2 calculations; however, we considered those warnings as false positives. The scripts used for each of the analyses are reported in Appendix 2.

Results - Experiment 1

The mean RTs and percentage error rates for the “same” targets are shown in Table 1 and the values of the test statistics from the analyses are shown in Tables 2 (latencies) and 3 (error rates).

Table 1 Mean decision latencies (reaction times (RTs), in milliseconds) and percentage error rates for “same” trials with Japanese participants in Experiment 1 (standard deviations in parentheses)

	One-character words		Two-character words	
	RT	%E	RT	%E
Homophone primes	454 (67)	6.1 (5)	464 (71)	6.4 (6)
Phonological unrelated	459 (65)	8.0 (6)	473 (62)	9.2 (6)
Priming	5	1.9	9	2.8

%E percentage error rate

The overall mean RT and error rate of the different trials were 481 ms and 3.6%, respectively

In the latency data, there was a significant Relatedness effect in all five analyses due to the fact that targets following homophonic primes (459 ms) were processed faster than targets following phonologically unrelated primes (466 ms). This effect had a modest size, and, even though it was significant in all five analyses, three of the four analyses for which power could be calculated were a bit short of the .80 power level (range: .652–.926). The main effect of Word length was also significant in all the analyses, reflecting the fact that one-character target words (456 ms) were processed faster than two-character target words (468 ms). This effect had a larger size and virtually all of the analyses had a power level above .80 (range: .786–.998). Although the effect of Relatedness was slightly larger for two-character words (9 ms) than for one-character words (5 ms), none of the analyses suggested an interaction, all $ps > 0.1$. Assuming an interaction of this size is real, all four of the analyses that allowed a power calculation were severely underpowered to detect it (range .109–.293).

There was also a main effect of Relatedness in all the error data analyses, reflecting the fact that the error rate was lower following homophonic primes (6.3%) than following phonologically unrelated primes (8.7%). Neither the main effect of Word length nor the interaction between Relatedness and Word length approached significance, all $ps > 0.3$. The main effect of Relatedness was the only effect with a power of at least .80 in the analyses that allowed a power calculation.

Results - Experiment 2

The mean RTs and proportion error rates for the “same” targets are shown in Table 4 and the values of the test statistics from the analyses are shown in Tables 5 (latencies) and 6 (error rates).

The latency data showed a significant Relatedness effect in all five analyses, reflecting the fact that targets following homophonic primes (560 ms) were processed faster than targets following phonologically unrelated primes (572 ms), an effect that was slightly larger than in Experiment 1. All of the analyses for which power could be calculated showed a power of at least .80. The main effect of Word length was significant in all the five analyses, reflecting the fact that one-character words were processed faster (560 ms) than two-character words (572 ms). This effect had a size comparable to that of the Relatedness effect, and all of the analyses had a power of at least .80. Although, again, there was a numerical tendency for a larger effect of Relatedness for two-character words (14 ms) than for one-character words (9 ms), none of the analyses suggested an interaction, all $ps > 0.1$. While the ANOVAs were severely

Table 2 Latency analysis results from the five analysis techniques for Experiment 1 (Japanese participants)

Analysis	Relatedness	Length	Interaction
F_s	$F(1, 34) = 6.90, MSE = 271, p = .013, \eta_p^2 = .169, \text{power} = .724$	$F(1, 34) = 25.57, MSE = 209, p < .001, \eta_p^2 = .429, \text{power} = .998$	$F(1, 34) = .73, MSE = 265, p = .401, \eta_p^2 = .021, \text{power} = .132$
F_i	$F(1, 116) = 6.71, MSE = 653, p = .011, \eta_p^2 = .055, \text{power} = .731$	$F(1, 116) = 10.85, MSE = 922, p = .001, \eta_p^2 = .086, \text{power} = .786$	$F(1, 116) = .98, MSE = 653, p = .324, \eta_p^2 = .008, \text{power} = .109$
LMM raw data	$\beta = -3.55, SE = 1.41, t = -2.51, p = .012, \text{semipartial } R^2 = .001, \text{power} = .652 (95\% \text{ CI } [.622, .682])$	$\beta = -6.12, SE = 1.90, t = -3.22, p = .002, \text{semipartial } R^2 = .003, \text{power} = .807 (95\% \text{ CI } [.781, .831])$	$\beta = 1.24, SE = 1.41, t = .88, p = .380, \text{semipartial } R^2 < .001, \text{power} = .293 (95\% \text{ CI } [.265, .322])$
LMM transformed data	$\beta = -.02, SE = .01, t = -3.73, p < .001, \text{semipartial } R^2 = .002, \text{power} = .926 (95\% \text{ CI } [.908, .942])$	$\beta = -.04, SE = .01, t = -4.10, p < .001, \text{semipartial } R^2 = .005, \text{power} = .957 (95\% \text{ CI } [.943, .969])$	$\beta = .01, SE = .01, t = 1.19, p = .235, \text{semipartial } R^2 < .001, \text{power} = .265 (95\% \text{ CI } [.238, .294])$
GLMM raw data*	$\beta = -4.23, SE = 1.26, z = -3.36, p < .001, \text{semipartial } R^2 = .001$	$\beta = -6.21, SE = 2.24, z = -2.77, p = .006, \text{semipartial } R^2 = .003$	$\beta = 1.76, SE = 1.26, z = 1.41, p = .160, \text{semipartial } R^2 < .001$

* Results from model restarted from apparent optimum (initial model failed to converge) Convergence warnings for semipartial R^2 calculations were considered false positives

Table 3 Error rate analysis results from the three analysis techniques for Experiment 1 (Japanese participants)

Analysis	Relatedness	Length	Interaction
F_s	$F(1, 34) = 10.05, MSE = .002, p = .003, \eta_p^2 = .228, \text{power} = .868$	$F(1, 34) = 1.02, MSE = .002, p = .321, \eta_p^2 = .029, \text{power} = .165$	$F(1, 34) = .26, MSE = .002, p = .616, \eta_p^2 = .007, \text{power} = .076$
F_i	$F(1, 116) = 8.73, MSE = .004, p = .004, \eta_p^2 = .070, \text{power} = .834$	$F(1, 116) = .82, MSE = .005, p = .369, \eta_p^2 = .007, \text{power} = .101$	$F(1, 116) = .26, MSE = .004, p = .615, \eta_p^2 = .002, \text{power} = .064$
GLMM	$\beta = .19, SE = .06, z = 3.18, p = .001, \text{semipartial } R^2 = .002$	$\beta = .05, SE = .07, z = .71, p = .476, \text{semipartial } R^2 < .001$	$\beta = -.04, SE = .06, z = -.66, p = .509, \text{semipartial } R^2 < .001$

underpowered to detect such an interaction, the power of the LMM analysis on the untransformed data did approach .80.⁷

In the error rate analysis, the main effect of Relatedness was significant in all three analyses, suggesting that related trials

(5.4%) produced fewer errors than unrelated trials (7.2%). Neither the main effect of Word length nor the interaction approached significance in any of the analyses, all $ps > 0.1$. The main effect of Relatedness in the F_i analysis had a power of over .80, whereas the power in the F_s analysis was .655.

⁷ A reviewer on a previous version of this article noted that the fact that, in both Experiment 1 and Experiment 2, phonological priming was numerically larger for two-character words than for one-character words would be consistent with the idea that longer stimuli may allow more time for phonological activation/priming to emerge. Although not central to the present research, that idea is certainly plausible. However, our results suggest that a very sensitive protocol would be required to substantiate that idea in the present paradigm, i.e., to obtain a significant Relatedness by Word length interaction. We conducted additional power analyses using G*Power software (for ANOVAs; Faul et al., 2009) and the simR package in R (for LMMs; Green & MacLeod, 2016) to determine the sample size that would be required to achieve a .80 power to detect an interaction of the size we found in Experiments 1 and 2 in the latencies. For Experiment 1, a sample size of 370 participants (power = .800), 1,360 items (power = .800), 294 participants (power = .807, 95% CI [.781, .831]), and 184 participants (power = .803, 95% CI [.777, .827]) would have been required for the $F_s, F_i, \text{LMM raw data}$ and $\text{LMM transformed data}$ analyses, respectively, to have achieved power levels of at least .80. For Experiment 2, a sample size of 488 participants (power = .801), 2,180 items (power = .801), 118 participants (power = .810, 95% CI [.784, .834]), and 70 participants (power = .807, 95% CI [.781, .831]) would have been required for the $F_s, F_i, \text{LMM raw data},$ and $\text{LMM transformed data}$ analyses, respectively, to have achieved power levels of at least .80. These sample sizes exceed, often by a large margin, the sample sizes typically used in cognitive psychology experiments.

Discussion

The results are fairly straightforward. There were small but significant phonological priming effects in the masked

Table 4 Mean decision latencies (reaction times (RTs), in milliseconds) and percentage error rates for “same” trials with Chinese participants in Experiment 2 (standard deviations in parentheses)

	One-character words		Two-character words	
	RT	%E	RT	%E
Homophone primes	556 (70)	5.6 (6)	565 (68)	5.2 (5)
Phonological unrelated	565 (67)	7.0 (8)	579 (64)	7.3 (8)
Priming	9	1.4	14	2.1

%E percentage error rate

The overall mean RT and error rate of the different targets were 592 ms and 3.1%, respectively

Table 5 Latency analysis results from the five analysis techniques for Experiment 2 (Chinese participants)

Analysis	Relatedness	Length	Interaction
F_s	$F(1, 52) = 9.40, MSE = 764, p = .003, \eta_p^2 = .153, \text{power} = .853$	$F(1, 52) = 12.98, MSE = 616, p < .001, \eta_p^2 = .200, \text{power} = .943$	$F(1, 52) = .83, MSE = 464, p = .366, \eta_p^2 = .016, \text{power} = .147$
F_i	$F(1, 116) = 8.51, MSE = 1007, p = .004, \eta_p^2 = .068, \text{power} = .823$	$F(1, 116) = 11.68, MSE = 765, p = .009, \eta_p^2 = .092, \text{power} = .818$	$F(1, 116) = .57, MSE = 1007, p = .451, \eta_p^2 = .005, \text{power} = .085$
LMM raw data	$\beta = -5.98, SE = 1.49, t = -4.01, p < .001, \text{semipartial } R^2 = .002, \text{power} = .992$ (95% CI [.984, .997])	$\beta = -6.05, SE = 1.73, t = -3.50, p < .001, \text{semipartial } R^2 = .002, \text{power} = .971$ (95% CI [.959, .981])	$\beta = 1.45, SE = 1.49, t = 0.97, p = .330, \text{semipartial } R^2 < .001, \text{power} = .563$ (95% CI [.532, .594])
LMM transformed data	$\beta = -.02, SE < .01, t = -5.57, p < .001, \text{semipartial } R^2 = .004, \text{power} = 1$ (95% CI [.995, 1])	$\beta = -.02, SE = .01, t = -4.45, p < .001, \text{semipartial } R^2 = .004, \text{power} = .998$ (95% CI [.993, 1])	$\beta = .01, SE < .01, t = 1.36, p = .174, \text{semipartial } R^2 < .001, \text{power} = .757$ (95% CI [.729, .783])
GLMM raw data*	$\beta = -6.19, SE = 1.33, z = -4.66, p < .001, \text{semipartial } R^2 = .002$	$\beta = -6.05, SE = 2.08, z = -2.91, p = .004, \text{semipartial } R^2 = .002$	$\beta = 1.78, SE = 1.34, z = 1.33, p = .184, \text{semipartial } R^2 < .001$

* Results from model restarted from apparent optimum (initial model failed to converge) Convergence warnings for semipartial R^2 calculations were considered false positives

priming SDT in both Japanese (using Kanji stimuli) and Chinese in both the latency and error data. These effects represent the first two examples of phonological priming in this task when the prime and (word) target are written in the same script. The more general conclusion, therefore, is that even in an experiment in which the primes and (word) targets involve the same orthography, phonological priming does emerge in an SDT.

General discussion

Two masked priming SDT experiments were carried out in order to evaluate whether it is possible to obtain phonological priming effects in that task when the prime and (word) target are written in the same script. As Kinoshita et al. (2018) had noted, there was no evidence that such was the case in the extant literature. The results of both Experiment 1 (with Japanese readers and Kanji stimuli) and Experiment 2 (with Chinese readers) indicate that the clear answer is “yes.” These results coupled with those of Lupker et al. (2015, 2018), who used cross-script primes and targets, solidifies the argument that phonological similarity does produce priming and, hence, that phonological information does play a role, in the SDT.

The fact that phonological priming effects have now been observed in the situation in which the prime and (word) target

are written in the same script has obvious implications for SDT experiments in alphabetic languages. Specifically, because orthographically related primes and targets in alphabetic languages are inevitably also phonologically related, it is, therefore, not possible to conclude that any presumed orthographic priming effects in those languages are completely orthographically based. When such effects are observed, phonology may very well be making some contribution and, unfortunately, it simply is not possible to determine how much of a contribution it might be making based on what we now know about the nature of the task.

Certainly, one could use the results of Lupker et al. (2015, 2018), and, to some degree, the present results, to argue that the contribution of phonology to priming effects in the SDT is not large. It is difficult, however, to know to what degree that argument can be extended to situations in which both phonology and orthography simultaneously contribute to the priming effect in the SDT (as they inevitably would in alphabetic script experiments). The reason is that, if those two factors are enhancing processing in the same way (i.e., at the same processing stage), they may be interacting in a way that makes the impact of phonology more potent. That is, the impact of phonology might be combining with that of orthography to produce what is referred to as an “overadditive” interaction (see, e.g., Pastizzo, Neely, & Tse, 2008, for a demonstration of an

Table 6 Error rate analysis results from the three analysis techniques used for Experiment 2 (Chinese participants)

Analysis	Relatedness	Length	Interaction
F_s	$F(1, 52) = 5.75, MSE = .003, p = .020, \eta_p^2 = .100, \text{power} = .655$	$F(1, 52) = .15, MSE = .002, p = .697, \eta_p^2 = .003, \text{power} = .067$	$F(1, 52) = .37, MSE = .003, p = .544, \eta_p^2 = .007, \text{power} = .091$
F_i	$F(1, 116) = 8.54, MSE = .002, p = .004, \eta_p^2 = .069, \text{power} = .828$	$F(1, 116) = .14, MSE = .002, p = .708, \eta_p^2 = .001, \text{power} = .032$	$F(1, 116) = .51, MSE = .002, p = .475, \eta_p^2 = .004, \text{power} = .078$
GLMM	$\beta = .16, SE = .05, z = 3.04, p = .002, \text{semipartial } R^2 = .001$	$\beta = -.04, SE = .05, z = -.77, p = .444, \text{semipartial } R^2 < .001$	$\beta = -.04, SE = .05, z = -.77, p = .441, \text{semipartial } R^2 < .001$

overadditive interaction involving orthographic and semantic priming in a lexical decision task).

The question of whether the impacts of orthography and phonology actually do combine in an overadditive fashion is, unfortunately, rather difficult to address experimentally. The way to do so would be to create conditions in which the priming provided by each factor could be evaluated independently and then to compare the sum of those priming effects to the effect produced by primes in which both factors are active simultaneously (as was done by Pastizzo et al., 2008). If overadditivity were to be observed, the implication would be that the impact of phonology (as a result of it combining with the impact of orthography) was somewhat stronger than that observed in the present data and the data of Lupker et al. (2015, 2018). How one could actually set up an experiment of this sort is not at all clear, however. That is, although the present experiments did allow us to evaluate the impact of phonological priming in the absence of orthographic influences, it is hard to envision a situation in which one could evaluate the impact of orthographic priming in the absence of phonological influences in virtually any language.

Another point to note when thinking about this question is that the phonological priming effects observed here were observed in logographic scripts. As noted above, a common assumption is that phonological coding based on logographic characters is more difficult (and, hence, slower) than phonological coding based on alphabetic characters (Perfetti et al., 1992). Given that SDT responding is typically quite rapid (as it was in the present experiments, particularly in Experiment 1), *a priori*, a clear expectation would have been that even if phonology does play a role in the SDT, one should not expect our logographic primes to produce large phonological priming effects. One could certainly argue, therefore, that when the primes (and targets) are written in an alphabetic script, a script that allows more rapid activation of phonology, the impact of phonology may be somewhat stronger than that suggested by the effects observed here.

A further point to note when considering this question is that, as noted above, although the Relatedness by Word length interaction was not significant in either experiment, in both experiments, the relatedness effect was slightly larger for the two-character words than for the one-character words (by 4 ms in Experiment 1 and by 5 ms in Experiment 2). However, as seen in Tables 2, 3, 5, and 6, these experiments had very little power to detect an interaction of this size. When considering the latency data, power estimates ranged from below .10 in the conventional items analyses to .757 in one of the LMM analyses in Experiment 2. Further, as reported in footnote 7, calculations of the minimum number of participants and items needed to have achieved a power of .80 would have been far beyond the numbers typically used in these types of experiments. Clearly, if this interaction is a theoretically important one, establishing its reality statistically will take great effort. However, if the trend in the present data (i.e., that longer targets produce larger phonological

priming effects) is a real one, that would be a further reason to suggest that the impact of phonology is likely larger in alphabetic languages (in which most words are much longer than the one- and two-character words used in the present experiments) than the effects observed here would suggest.

One final point to note concerning the sizes of these priming effects is that, unlike in most SDT experiments, the reference stimuli and targets in the present experiments were physically identical. In alphabetic language SDT experiments (e.g., Kinoshita et al., 2018; Kinoshita & Norris, 2009), the standard manipulation is to present the reference stimulus in one case (e.g., lowercase – face) and the target in the other (e.g., uppercase – FACE). Doing so prevents participants from carrying out the matching process based on low-level featural similarity and, hence, potentially reducing the size of the priming effect from higher level (e.g., orthographic, phonological) factors. Unfortunately, the same could not be done here because both Chinese and Kanji characters can only be written in one case. As such, although it's not possible to determine whether participants were able to use a feature-matching strategy in the present experiments, to the extent that they were able to do so the impact would have been to reduce the sizes of the priming effects. This fact also supports the idea that the impact of phonological priming in the SDT in alphabetic languages (in which the reference stimulus and target are presented in different cases) is probably larger than the effect sizes reported here might suggest.

The basic conclusion that the present data offer, therefore, is simply that even when using same-script primes and (word) targets in a masked priming SDT, the overall priming effect is likely some combination of the impacts of orthography and phonology. Hence, priming effects in such experiments cannot be assumed to provide an uncontaminated view of the orthographic coding process. Let us be clear, however, that the argument is not that these effects are purely phonologically based or even that phonology is the major player in producing priming effects in the SDT. Orthographic processing very likely plays a more central role in that process, which implies that the masked priming SDT should certainly be used as one of multiple tools in investigations of models of orthographic coding. Nevertheless, one should keep in mind that at least some component of the priming is likely phonologically based when interpreting the data from masked priming SDT experiments.

Author notes This research was partially supported by Natural Sciences and Engineering Research Council of Canada Grant A6333 to Stephen J. Lupker and a grant from the Japan Society for the Promotion of Science (JSPS) to Mariko Nakayama. Masahiro Yoshihara is now at the Center for Japanese-Language Testing, The Japan Foundation, Yotsuya, Shinjuku-ku, Tokyo, 4-3, Japan. We would like to thank Xue Junyi and Chen Huan for their assistance in data collection and analysis.

Data availability statement The datasets generated during and/or analyzed in the current research are available in the Open Science Framework repository, <https://osf.io/ayz9b/>.

Appendix 1

Word stimuli used in Experiment 1

These stimuli served as word targets in Experiment 1. HP = homophone prime, UP = unrelated prime.

Target	HP	UP
弧	個	念
姻	員	部
賊	屬	級
錠	冗	乙
慨	害	線
娛	碁	廊
汽	奇	垂
識	式	案
穀	酷	邪
峽	況	錄
貝	階	涉
塔	糖	脈
栽	才	屈
軌	規	容
晚	番	宅
菌	禁	略
聖	製	輸
歡	監	福
句	區	以
購	抗	督
曜	洋	閣
宙	駐	肉
簡	官	稅
項	校	予
服	副	審
王	忖	義

Target	HP	UP
醜	發行	死去
酒	主演	兩面
皆	解禁	倉庫
開	海岸	打診
朝	調停	友達
有	優秀	鑑定
解	回答	意向
恐	強行	三墨
怪	階段	紙面
辭	事例	產地
改	解明	稅率
配	敗戰	文字
退	對談	生態
扶	不要	通行
修	收支	兩方
私	指摘	活動
仕	使用	軍事
脂	志望	切手
酵	控訴	次男
強	競合	指紋
學	樂器	視線
後	校舍	授賞
改	回收	優先
近	禁止	表現
海	會場	部門
規	企画	相談

Target	HP	UP
票	評	族
史	資	制
券	県	約
英	衛	態
韻	院	府
殉	準	処
吟	銀	億
迅	陣	令
獵	兩	檢
囚	週	般
閱	悅	寸
錯	策	電
胃	医	援
菓	貨	普
庶	署	灣
腸	帳	魅
詞	師	佐
淨	条	融
啓	刑	密
往	央	債
誕	単	給
威	維	途
豪	号	愛
純	順	邦
郡	軍	百
劑	材	批
章	賞	盟
刊	完	紀
逮	隊	欧
腦	農	庁
宴	演	料
段	談	職
課	可	護

Target	HP	UP
商店	焦点	課長
原稿	現行	男子
機器	危機	認識
部隊	舞台	記念
尊称	損傷	規律
豆乳	投入	實質
秘訣	否決	人工
破門	波紋	震災
閉口	並行	終盤
恒例	高齡	給付
想起	早期	首都
腫瘍	主要	登場
輕傷	繼承	地盤
窮狀	球場	壓縮
等級	投球	目玉
採集	最終	憲法
陶器	冬季	新顔
刺殺	視察	会期
良心	兩親	輸送
磁氣	時期	派遣
水深	推進	関心
下降	加工	運賃
包装	放送	反発
歌詞	菓子	脅迫
注視	中止	範圍
帽子	防止	對話
開店	回轉	休暇
高層	構想	処分
教科	強化	行動
專攻	選考	座長
後援	公演	装置
視点	支店	納得
意思	医師	特別

Appendix 2

Scripts

Analysis	Latency	Errors
F _s	RT_bysj_aov = aov(RT ~ Relatedness*Length*Group + Error(Subject/(Relatedness*Length)), aggregate(RT ~ Relatedness+Length+Group, data_without_errors, mean)) eta_sq(RT_bysj_aov, partial = T)	ER_bysj_aov = aov(accuracy ~ Relatedness*Length*Group + Error(Subject/(Relatedness*Length)), aggregate(accuracy ~ Relatedness+Length+Group, data_with_errors, mean)) eta_sq(ER_bysj_aov, partial = T)
F _i	RT_byitem_aov = aov(RT ~ Relatedness*Length*Set + Error(Target/Relatedness), aggregate(RT ~ Relatedness+Length+Set, data_without_errors, mean)) eta_sq(RT_byitem_aov, partial = T)	ER_byitem_aov = aov(accuracy ~ Relatedness*Length*Set + Error(Target/Relatedness), aggregate(accuracy ~ Relatedness+Length+Set, data_with_errors, mean)) eta_sq(ER_byitem_aov, partial = T)
LMM raw data	RT_lmm_raw = lmer(RT ~ Relatedness*Length*Group + (1 Subject) + (1 Target), data_without_errors) r2RT_lmm_raw = r2beta(RT_lmm_raw, method='nsj', partial=T) powerSim(RT_lmm_raw, fcompare(~ Length*Group, "lr"), nsim = 1000) powerSim(RT_lmm_raw, fcompare(~ Relatedness*Group, "lr"), nsim = 1000) powerSim(RT_lmm_raw, fcompare(~ Relatedness + Length + Group + Relatedness:Group + Length:Group, "lr"), nsim = 1000)	
LMM transformed data	RT_lmm_tra = lmer(-1000/RT ~ Relatedness*Length*Group + (1 Subject) + (1 Target), data_without_errors) r2RT_lmm_tra = r2beta(RT_lmm_tra, method='nsj', partial=T) powerSim(RT_lmm_tra, fcompare(~ Length*Group, "lr"), nsim = 1000) powerSim(RT_lmm_tra, fcompare(~ Relatedness*Group, "lr"), nsim = 1000) powerSim(RT_lmm_tra, fcompare(~ Relatedness + Length + Group + Relatedness:Group + Length:Group, "lr"), nsim = 1000)	
GLMM raw data (including syntax for restarting)	RT_glmm = glmer(RT ~ Relatedness*Length*Group + (1 Subject) + (1 Target), data_without_errors, family = Gamma(link = "identity"), control=glmerControl(optCtrl=list(maxfun=1e6))) RT_glmm.restart = update(RT_glmm, start= getME(RT_glmm, c("theta", "fixef"))) r2RT_glmm.restart = r2beta(RT_glmm.restart, method='nsj', partial=T)	ER_glmm = glmer(accuracy ~ Relatedness*Length*Group + (1 Subject) + (1 Target), data_with_errors, family = binomial, control=glmerControl(optCtrl=list(maxfun=1e6))) r2ER_glmm = r2beta(ER_glmm, method='nsj', partial=T)

References

- Amano, N. & Kondo, K. (2003). NTT detabesu shirizu: Nihongo no goi-tokusei, Dai 2-ki CD-ROM-ban [NTT database series: Lexical properties of Japanese, vol. 2, CD-ROM version]. Tokyo: Sanseido.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R. Processing*. Cambridge: Cambridge University Press. Retrieved from <http://www.sfs.uni-tuebingen.de/~hbaayen/publications/baayenCUPstats.pdf>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bolker, B. (2020). *GLMM FAQ*. Retrieved from <https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html>

- Brysbaert, M. & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1(1), 1–20. <https://doi.org/10.5334/joc.10>
- Cai, Q. & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, 5(6), e10729. <https://doi.org/10.1371/journal.pone.0010729>
- Castles, A., Davis, C., & Letcher, T. (1999). Neighbourhood effects on masked form priming in developing readers. *Language & Cognitive Processes*, 14(2), 201–224. <https://doi.org/10.1080/016909699386347>
- Coltheart, M., Davelaar, E., Jonasson, J. F., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, N.J.: Erlbaum.
- Davis, C.J. (2010). The spatial coding model of visual word identification. *Psychological Review*, 117(3), 713–758. <https://doi.org/10.1037/a0019738>
- Davis, C. J. & Lupker, S. J. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance*, 32(3), 668–687. <https://doi.org/10.1037/0096-1523.32.3.668>
- Duñabeitia, J. A., Kinoshita, S., Carreiras, M., & Norris, D. (2011). Is morpho-orthographic decomposition purely orthographic? Evidence from masked priming in the same-different task. *Language & Cognitive Processes*, 26(4–6), 509–529. <https://doi.org/10.1080/01690965.2010.499215>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Ferrand, L. & Grainger, J. (1992). Phonology and orthography in visual word recognition: Evidence from masked non-word priming. *The Quarterly Journal of Experimental Psychology Section A*, 45(3), 353–372. <https://doi.org/10.1080/02724989208250619>
- Ferrand, L. & Grainger, J. (1993). The time course of orthographic and phonological code activation in the early phases of visual word recognition. *Bulletin of the Psychonomic Society*, 31(2), 119–122. <https://doi.org/10.3758/BF03334157>
- Forster, K. I. & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(4), 680–698. <https://doi.org/10.1037/0278-7393.10.4.680>
- Forster, K. I., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language*, 30(1), 1–25.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35(1), 116–124. <https://doi.org/10.3758/bf03195503>
- 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>
- García-Orza, J., Perea, M., & Estudillo, A. (2011). Masked transposition effects for simple versus complex nonalphanumeric objects. *Attention, Perception, & Psychophysics*, 73(8), 2573–2582. <https://doi.org/10.3758/s13414-011-0206-7>
- García-Orza, J., Perea, M., & Muñoz, S. (2010). Are transposition effects specific to letters? *Quarterly Journal of Experimental Psychology*, 63(8), 1603–1618. <https://doi.org/10.1080/17470210903474278>
- Gómez, P., Perea, M. & Ratcliff, R. (2008). The overlap model: A model of letter position coding. *Psychological Review*, 115(3), 577–601. <https://doi.org/10.1037/a0012667>
- Grainger, J., Granier, J.P., Farioli, F., Van Assche, E. & Van Heuven, W. (2006). Letter position information and printed word perception: The relative-position priming constraint. *Journal of Experimental Psychology: Human Perception and Performance*, 32(4), 865–884. <https://doi.org/10.1037/0096-1523.32.4.865>
- Grainger, J., & Van Heuven, W. J. B. (2003). Modeling letter position coding in printed word perception. In P. Bonin (Ed.), *The mental lexicon* (pp. 1–23). New York: Nova Science.
- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. <https://doi.org/10.1111/2041-210X.12504>
- Hino, Y., Kusunose, Y., Lupker, S. J. & Jared, D. (2013). The processing advantage and disadvantage for homophones in lexical decision tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(2), 529–551. <https://doi.org/10.1037/a0029122>
- 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>
- Jaeger, B. C., Edwards, L. J., Das, K., & Sen, P. K. (2017). An R 2 statistic for fixed effects in the generalized linear mixed model. *Journal of Applied Statistics*, 44(6), 1086–1105. <https://doi.org/10.1080/02664763.2016.1193725>
- Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth’s R2GLMM to random slopes models. *Methods in Ecology and Evolution*, 5(9), 944–946. <https://doi.org/10.1111/2041-210X.12225>
- Kinoshita, S. (2000). The left-to-right nature of the masked onset priming effect in naming. *Psychonomic Bulletin & Review*, 7(1), 133–141. <https://doi.org/10.3758/bf03210732>
- Kinoshita, S., Gayed, M., & Norris, D. (2018). Orthographic and phonological priming effects in the same-different task. *Journal of Experimental Psychology: Human Perception and Performance*, 44(11), 1661–1671. <https://doi.org/10.1037/xhp0000548>
- Kinoshita, S. & Norris, D. (2009). Transposed-letter priming of prelexical orthographic representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1), 1–18. <https://doi.org/10.1037/a0014277>
- Kinoshita, S. & Norris, D. (2010). Masked priming effect reflects evidence accumulated by the prime. *Quarterly Journal of Experimental Psychology*, 63(1), 194–204. <https://doi.org/10.1080/17470210902957174>
- Kinoshita, S., Schubert, T. & Verdonschot, R. G. (2019). Allograph priming is based on abstract letter identities: Evidence from Japanese kana. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(1), 183–190. <https://doi.org/10.1037/xlm0000563>
- 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>
- Levy, R. (2014). Using R formulae to test for main effects in the presence of higher-order interactions. *ArXiv Preprint ArXiv:1405.2094*. Retrieved from <https://arxiv.org/pdf/1405.2094.pdf>
- Li, X., Rayner, K., & Cave, K. R. (2009). On the segmentation of Chinese words during reading. *Cognitive Psychology*, 58(4), 525–552. <https://doi.org/10.1016/j.cogpsych.2009.02.003>
- Lo, S. & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, 30(6), 1171. <https://doi.org/10.3389/fpsyg.2015.01171>
- Lüdtke, D. (2020). sjstats: Statistical functions for regression models (version 0.18.0). <https://doi.org/10.5281/zenodo.1284472>
- Lupker, S. J., Nakayama, M., & Perea, M. (2015). Is there phonologically based priming in the same-different task? Evidence from Japanese-English bilinguals. *Journal of Experimental Psychology: Human Perception and Performance*, 41(5), 1281–1299. <https://doi.org/10.1037/xhp0000087>

- Lupker, S. J., Nakayama, M., & Yoshihara, M. (2018). Phonologically-based priming in the same-different task with L1 readers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(8), 1317–1324. <https://doi.org/10.1037/xlm0000515>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Norris, D. & Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General*, 137(3), 434–455. <https://doi.org/10.1037/a0012799>
- Norris, D. & Kinoshita, S. (2012). Reading through a noisy channel: Why there's nothing special about the perception of orthography. *Psychological Review*, 119(3), 517–545. <https://doi.org/10.1037/a0028450>
- Norris, D., Kinoshita, S. & Van Caasteren, M. (2010). A stimulus sampling theory of letter identity and order. *Journal of Memory and Language*, 62(3), 254–271. <https://doi.org/10.1016/j.jml.2009.11.002>
- Pastizzo, M. J., Neely, J. H., & Tse, C. S. (2008). With a letter-searched prime, *boat* primes *float* but *swim* and *coat* don't: Further evidence for automatic semantic activation. *Psychonomic Bulletin & Review*, 15(4), 845–849. <https://doi.org/10.3758/PBR.15.4.845>
- Perea, M. & Lupker, S. J. (2003). Transposed-letter confusability effects in masked form priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: The state of the art* (pp. 97–120). Philadelphia, PA: Psychology Press. <https://doi.org/10.4324/9780203502846>
- Perea, M. & Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language*, 51(2), 231–246. <https://doi.org/10.1016/j.jml.2004.05.005>
- Perea, M., Moret-Tatay, C., & Carreiras, M. (2011). Facilitation vs. inhibition in the masked priming same-different matching task. *Quarterly Journal of Experimental Psychology*, 64(10), 2065–2079. <https://doi.org/10.1080/17470218.2011.582131>
- Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychological Review*, 112(1), 43–59. <https://doi.org/10.1037/0033-295X.112.1.43>
- Perfetti, C. A. & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(1), 101–118. <https://doi.org/10.1037/0278-7393.24.1.101>
- Perfetti, C. A. & Zhang, S. (1995). Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(1), 24–33. <https://doi.org/10.1037/0278-7393.21.1.24>
- Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a “Universal” phonological principle. *Advances in Psychology*, 94, 227–248. [https://doi.org/10.1016/S0166-4115\(08\)62798-3](https://doi.org/10.1016/S0166-4115(08)62798-3)
- Pollatsek, A. & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(3), 785–794. <https://doi.org/10.1037/0278-7393.21.3.785>
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from <https://cran.r-project.org/doc/manuals/r-release/R-ints.html>
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Pittsburgh, PA: Psychology Software Tools.
- Schoonbaert, S., & Grainger, J. (2004). Letter position coding in printed word perception: Effects of repeated and transposed letters. *Language & Cognitive Processes*, 19(3), 333–367. <https://doi.org/10.1080/01690960344000198>
- Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. In D. H. Spieler & E. Schumacher (Eds.), *New methods in cognitive psychology* (pp. 4–31). Psychology Press. <https://doi.org/10.4324/9780429318405-2>
- Tamaoka, K. (2007). Rebounding activation caused by lexical homophony in the processing of Japanese two-kanji compound words. *Reading and Writing*, 20(5), 413–439. <https://doi.org/10.1007/s11145-006-9036-8>
- Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review*, 8(2), 221–243. <https://doi.org/10.3758/BF03196158>
- Whitney, C. & Marton, Y. (2013). *The SERIOL2 Model of Orthographic Processing*. Retrieved from http://www.researchgate.net/publication/237065841_The_SERIOL2_Model_of_Orthographic_Processing/file/5046351b204501e42a.pdf
- Wydell, T. N., Patterson, K. E. & Humphreys, G. W. (1993). Phonologically mediated access to meaning for Kanji: Is a rose 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>
- Yang, H., Chen, J., Spinelli, G., & Lupker, S. J. (2019). The impact of text orientation on form priming effects in four-character Chinese words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(8), 1511–1526. <https://doi.org/10.1037/xlm0000655>
- Zhou, X. & Marslen-Wilson, W. (1999). Phonology, orthography, and semantic activation in reading Chinese. *Journal of Memory and Language*, 41(4), 579–606. <https://doi.org/10.1006/jmla.1999.2663>
- Zhou, X. & Marslen-Wilson, W. (2000). The relative time course of semantic and phonological activation in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1245–1265. <https://doi.org/10.1037/0278-7393.26.5.1245>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.