Journal of Experimental Psychology:
Learning, Memory, and Cognition
2009, Vol. 35, No. 1, 238–246

0278-7393/09/$12.00 DOI: 10.1037/a0014361

The Cross-Script Length Effect: Further Evidence Challenging PDP Models of Reading Aloud

Kathleen Rastle
Royal Holloway, University of London

Jelena Havelka
University of Leeds

Taeko N. Wydell
Brunel University

Max Coltheart
Macquarie University

Derek Besner
University of Waterloo

The interaction between length and lexical status is one of the key findings used in support of models of reading aloud that postulate a serial process in the orthography-to-phonology translation (B. S. Weekes, 1997). However, proponents of parallel models argue that this effect arises in peripheral visual or articulatory processes. The authors addressed this possibility using the special characteristics of the Serbian and Japanese writing systems. Experiment 1 examined length effects in Serbian when participants were biased to interpret phonologically bivalent stimuli in the alphabet in which they are words or in the alphabet in which they are nonwords (i.e., the visual characteristics of stimuli were held constant across lexical status). Experiment 2 examined length effects in Japanese kana when words were presented in the kana script in which they usually appear or in the script in which they do not normally appear (i.e., the phonological characteristics of stimuli were held constant across lexical status). Results in both cases showed a larger length effect when stimuli were treated as nonwords and thus offered strong support to models of reading aloud that postulate a serial component.

Keywords: reading aloud, serial processing, length effects, cross-linguistic, PDP models

The past 20 years has seen significant progress in our understanding of the mental processes involved in reading aloud. Several competing computational models are now available, among them, the DRC model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), the CDP+ model (Perry, Ziegler, & Zorzi, 2007), and the family of parallel-distributed-processing (PDP) "triangle" models (Harm & Seidenberg, 2004; Plaut, McClelland, Seidenberg, & Patterson, 1996). Though each of these models is able to simulate most of the benchmark phenomena within the domain, they all make fundamentally different claims about the mechanisms underlying the transformation of orthography to phonology. For this reason, it has become increasingly important to direct our research efforts toward problems that offer the potential for adjudicating between the models.

One such problem concerns the nature of processing in the reading aloud system. In particular, whereas the PDP models of reading aloud (Harm & Seidenberg, 2004; Plaut et al., 1996) make the a priori claim that orthographic input representations are translated to phonological output representations solely through the use of parallel processing, the DRC (Coltheart et al., 2001) and CDP+ (Perry et al., 2007) models posit that this translation is accomplished through the use of a combination of parallel and serial processes. Though these latter models have important differences, both are dual-route theories in which a lexical route computes the phonological code in a fully parallel manner and in which a nonlexical route assembles the phonological code serially, from left to right, across the letter string.

Several phenomena have been offered as evidence for serial processing in reading aloud. These include the position of irregularity effect (Coltheart & Rastle, 1994; Rastle & Coltheart, 1999; Roberts, Rastle, Coltheart, & Besner, 2003), the whammy effect (Rastle & Coltheart, 1998), the position-sensitive Stroop effect (Coltheart, Woollams, Kinoshita, & Perry, 1999), the position of bivalence effect (Havelka & Rastle, 2005), and the interaction

Kathleen Rastle, Department of Psychology, Royal Holloway, University of London, Egham, Surrey, United Kingdom; Jelena Havelka, Institute of Psychological Sciences, University of Leeds, Leeds, West Yorkshire, United Kingdom; Taeko N. Wydell, Department of Psychology, Brunel University, Uxbridge, Middlesex, United Kingdom; Max Coltheart, Macquarie Centre for Cognitive Science, Macquarie University, Sydney, New South Wales, Australia; Derek Besner, Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada.

This work was supported by Economic and Social Research Council Grant RES-000-22-1875. Experiment 2 was also supported by a travel grant to Taeko N. Wydell from Macquarie University. We are grateful to Lyndsey Nickels and Aleksandar Kostíc for helping us to recruit our Japanese and Serbian participants and to Lisa Yen and James Milward for technical assistance.

Correspondence concerning this article should be addressed to Kathleen Rastle, Department of Psychology, Royal Holloway, University of London, Egham, Surrey TW20 0EX, United Kingdom. E-mail: Kathy.Rastle@rhul.ac.uk

between length and lexical status (Weekes, 1997; see Rastle & Coltheart, 2006, for a review of these phenomena). However, despite this apparent abundance of evidence, it has been difficult to achieve consensus on the parallel versus serial question, for the reason that proponents of parallel models have been able to lodge reasonably persuasive alternative accounts that do not involve serial processing.

Alternative Accounts of Apparent Serial Effects

Proponents of parallel models have offered two kinds of explanations for effects on reading aloud that appear to reflect serial processing. One explanation is that these effects reflect not serial processing but rather the influence of some other factor inadvertently confounded with the serial manipulation of interest. One example of this problem concerns the interpretation of the position of irregularity effect (e.g., Rastle & Coltheart, 1999). The position of irregularity effect refers to the finding that the regularity effect on reading aloud is modulated by the position in the word at which the irregularity occurs, such that the cost of irregularity declines as the position of irregularity moves from left to right. Rastle and Coltheart claimed that this effect is inconsistent with any model that accomplishes the orthography-to-phonology transformation in parallel. However, this claim was undermined by Zorzi (2000), who demonstrated that the effect could be simulated by a fully parallel model, as a result of the fact that the position of irregularity factor in Rastle and Coltheart's study had been confounded with grapheme–phoneme consistency. Stimuli containing early irregularities (e.g., ***ch**ef*) were more inconsistent than stimuli containing later irregularities (e.g., *gl**ow***), and the parallel model was sensitive to this difference. Though this specific issue of grapheme–phoneme consistency and the position of irregularity effect appears to have been resolved (Roberts et al., 2003), such unintentional confounds are always a risk when comparisons across different targets are used (see also Seidenberg, Petersen, MacDonald, & Plaut, 1996, for a similar argument concerning the pseudohomophone advantage in reading aloud).

The other explanation lodged by proponents of parallel models is that serial effects do occur, but that they arise in peripheral components of the reading aloud system (such as early visual encoding and speech motor execution) that are outside the scope of implemented models (see, e.g., Kawamoto, Kello, Jones, & Bame, 1998; Plaut, 1999; Seidenberg & Plaut, 1998). One example of this kind of argument was developed by Kawamoto et al. (1998) in relation to the position of irregularity effect. Specifically, they claimed that the orthography-to-phonology translation occurs in parallel but that the position of irregularity effect arises because of the serial nature of articulation. By their account, a cost of irregularity is observed for words with initial irregularities (e.g., ***ch**ef*) because articulation (and thus the emission of acoustic energy) cannot begin until these phonemes are computed, whereas no cost of irregularity is apparent for words with late irregularities (e.g., *gl**ow***) because these phonemes can be computed during articulation of the preceding phonemes. It is because articulation is a serial process that no irregularity disadvantage is observed for stimuli with late irregularities (but see Rastle, Harrington, Coltheart, & Palethorpe, 2000). The argument is that because implemented models do not deal with articulatory processes, they cannot be expected to simulate these kinds of effects.

Length Effects and Serial Processing

The research described in this article contributes to a resolution of these issues. We studied the interaction between length and lexical status using a cross-linguistic approach that enabled us to address the alternative explanations put forward by proponents of parallel models. The interaction between length and lexical status refers to the finding that reading aloud latency increases as the number of letters in a stimulus increases but that this increase is much larger for nonwords than it is for words (Weekes, 1997; see also Ziegler, Perry, Jacobs, & Braun, 2001). The DRC and CDP+ models explain this interaction as a consequence of the serially operating nonlexical procedure (Coltheart et al., 2001; Perry et al., 2007). The reason that this effect is so strong for nonwords is that the nonlexical procedure makes a far greater contribution to the reading aloud of nonwords than it does to the reading aloud of words in these models.

Although this interaction has been one of the key findings used in support of models that incorporate a serial component, the interaction's significance can be challenged by the alternative explanations described above. It is not hard to imagine that small length effects could arise due to the serial encoding of letters (e.g., Whitney, 2001) or to serial aspects of articulation (e.g., Kawamoto et al., 1998). Indeed, Seidenberg and Plaut (1998) made precisely this argument in their analysis of item-specific variance in reading aloud data: "The residual effects of length are a reminder that there are aspects of word recognition and pronunciation beyond the scope of implemented models" (p. 235). Of course, such length effects would apply to words and nonwords alike—quite unlike the pattern of data reported by Weekes (1997). However, if there were also some uncontrolled difference between word and nonword stimuli, these small length effects could become exaggerated for nonwords.

What might these uncontrolled differences be? Because the manipulation of lexical status involves a between-items comparison, any number of important differences could characterize these groups of stimuli. Nonwords constitute unfamiliar visual and articulatory patterns and thus could be characterized by lower bigram/trigram or biphone/triphone probabilities, lesser body or rime frequencies, and/or lesser letter frequencies than are words (just to name a few differences; see Seidenberg et al., 1996, for a detailed discussion of this issue). If the reading system were characterized by serial encoding of letters, for example, it is not unreasonable to believe that this serial encoding process could be modulated by a variable such as letter frequency; stimuli comprising less frequent letters would be encoded more slowly. Such speculations about potential differences between word and nonword stimuli are made all the more serious in light of the fact that one fully parallel model (Zorzi, Houghton, & Butterworth, 1998) has successfully simulated the interaction between length and lexical status, both in the context of the stimuli used by Weekes (1997; see Zorzi, 2000, p. 855)[1] and in

[1] Though Zorzi (2000) argued that the model developed by Zorzi et al. (1998) produced a significant interaction between length and lexical status, the simulations with this model published by Coltheart et al. (2001) showed that the interaction failed to reach significance. The discrepancy apparently relates to a single item that Zorzi (2000) classified as an outlier but that Coltheart et al. did not (see Perry & Ziegler, 2002, p. 994).

the context of the English stimuli used by Ziegler et al. (2001; see Perry & Ziegler, 2002, Figure 1).

One interesting approach to resolving some of these issues was taken by Perry and Ziegler (2002; see also Ziegler et al., 2001). In the course of an investigation on the use of various grain sizes in reading different languages, Ziegler et al. (2001) had designed sets of German and English words and nonwords that varied in letter length. Critical to their manipulation was the fact that the German and English stimuli were cognates (i.e., stimuli that look and sound identical and, in the case of words, mean the same thing; e.g., *zoo–Zoo, sand–Sand*). Ziegler et al. (2001) described these stimuli as "literally identical" (p. 383) or, in cases in which this could not be achieved, as "orthographically and phonologically . . . similar as possible" (p. 381). Ziegler et al. replicated the interaction between length and lexical status in each language. However, their critical finding was that the length effect overall (collapsed across lexical status) was larger in German than in English. Perry and Ziegler (2002) argued that this Length × Language interaction rules out the "peripheral processes" explanation of the length effect lodged by Seidenberg and Plaut (1998) because the peripheral processes required for performing the task were matched extremely well (through the use of cognates) across the manipulation of language.

However, an inspection of the stimuli used by Ziegler et al. (2001), which were published in Perry and Ziegler (2002), reveals that the German and English items were some considerable distance from being "identical." Indeed, only 23/80 words and 37/80 nonwords used by Ziegler et al. were orthographically identical and only 11/80 words and 8/80 nonwords were phonologically identical.[2] To make matters worse, there are numerous examples of cases in which there are substantial orthographic and phonological differences across the English–German language manipulation among the word stimuli (e.g., *toe–zeh, hay–heu, cloth–kleid, cross–kreuz, change–schlaf, fierce–feucht*) and the nonword stimuli (e.g., *bry–pei, spond–spaut, gladge–klackt, dright–drecht*). In the light of these shortcomings, we believe, it would be premature to rule out the peripheral processes explanation of the length effect on the basis of Ziegler et al.'s findings.

The Present Experiments

It should be clear that interpreting the interaction between length and lexical status as evidence for serial processing is complicated by the fact that there are numerous potentially important orthographic and phonological differences between words and nonwords. The approach we took in the following experiments was thus to examine the interaction between length and lexical status in a situation in which the orthographic and phonological properties of words and nonwords were better controlled than had previously been possible. The special properties of the Serbian and Japanese writing systems provided us with precisely this situation. Use of the Serbian writing system allowed us to hold orthography constant across the manipulation of lexical status, such that words and nonwords comprised identical orthographic forms (Experiment 1). Similarly, use of the Japanese writing system allowed us to hold phonology constant across the manipulation of lexical status, such that words and nonwords comprised identical phonological forms (Experiment 2). Though a length effect could still arise due to visual or articulatory differences across stimuli of various lengths, there would be no possibility with this design to ascribe a larger length effect for nonwords to such differences. Thus, if an interaction between length and lexical status persisted under these conditions, it would be very difficult to explain in the context of a fully parallel model.

Experiment 1

Experiment 1 investigated the interaction between length and lexical status in Serbian. The Serbian language is transcribed into two alphabets, Cyrillic and Roman, both of which are characterized by perfect print-to-sound consistency. Each of these alphabets comprises three types of letters: (a) unique letters that occur in only one of the alphabets; (b) common letters that occur in both alphabets and map onto the same phonemes in each; and (c) ambiguous letters that occur in both alphabets but map onto different phonemes in each (see Figure 1). These types of letters can be used to construct (a) unique words comprising unique and common letters that can be read in only one alphabet and (b) bivalent words comprising ambiguous and common letters that have two possible pronunciations depending upon the alphabet in which the string is read. Bivalent strings are typically meaningful in only one alphabet; they are pronounceable nonwords (with a different pronunciation of the ambiguous letters) in the other. For example, the stimulus *PAHA* is a word meaning "wound" in Cyrillic and is pronounced /rana/, whereas it is a nonword in Roman and is pronounced /paha/.

Havelka and Rastle (2005) described a dual-route theory of Serbian reading aloud, the key features of which included (a) a single orthographic lexicon comprising all known Roman and Cyrillic words and (b) a single nonlexical translation procedure comprising spelling-to-sound rules for both Cyrillic and Roman letters (or, in the case of ambiguous letters, Cyrillic and Roman interpretations of these letters). These features of the theory are consistent with recent findings on bilingual word recognition and reading aloud (e.g., Brysbaert, 2003; Dijkstra & Van Heuven, 2002; Jared & Kroll, 2001; van Wijnendaele & Brysbaert, 2002). According to this theory, two pronunciations are computed for bivalent words (one word pronunciation and one nonword pronunciation), and the resulting conflict gives rise to the phonological bivalence effect (i.e., slower reading aloud of bivalent words than of unique words; Lukatela, Lukatela, Carello, & Turvey, 1999). This conflict (as measured by the size of the bivalence effect) can be reduced substantially by (a) presenting Roman and Cyrillic stimuli in separate blocks of trials, (b) including Roman or Cyrillic unique fillers with target bivalent stimuli as appropriate, and (c) instructing participants that they will be presented with Roman or Cyrillic stimuli only (Havelka & Rastle, 2005).

Experiment 1 thus studied participants' reading aloud of one set of bivalent targets varied on length, half of which were words in Roman and half of which were words in Cyrillic. This full set of targets was presented to separate groups of Roman and Cyrillic readers, and the targets were mixed as appropriate with unique Roman or Cyrillic fillers. In this way, we were able to examine the size of the length effect across word and nonword contexts, in a

[2] These phonological judgments were made by two native speakers of German.

Uniquely Cyrillic letters	Common letters	Uniquely Roman letters
Б Ц Ч Ћ Д	A E O J K M T	Č Ć D Đ F
Ђ Џ Ф Г Х		G I L N R
И Л Љ Њ П	H P C B	S Š U V LJ
Ш У З Ж	Bivalent letters	Z DŽ Ž NJ

Figure 1. Letters of the Roman and Cyrillic alphabets.

situation in which exactly the same orthographic strings were used across the manipulation of lexical status. Models that postulate a serial nonlexical component predict that the interaction between length and lexical status will be maintained under these strict conditions. Models that operate solely in parallel do not.

Method

Participants. Fifty-two students from the University of Belgrade participated in the experiment as part of a course requirement. Half were assigned to the Cyrillic version of the experiment, and half were assigned to the Roman version of the experiment. Participants had normal or corrected-to-normal vision, were native speakers of Serbian, and were free from any reading impairments. All participants were very familiar with both alphabets.

Materials and apparatus. Target stimuli comprised 60 bivalent items. Thirty of them were words when interpreted in the Roman alphabet and were pronounceable nonwords when interpreted in the Cyrillic alphabet; the other 30 were words when interpreted in the Cyrillic alphabet and were pronounceable nonwords when interpreted in the Roman alphabet. Thus, each of these targets was read aloud as a word or as a nonword by our participants, depending on the version of the experiment (Roman or Cyrillic) to which they were assigned.

Each set of 30 bivalent targets consisted of 10 items in each of three length conditions (4, 5, or 6 letters). The sets of bivalent targets were matched on frequency across the three levels of length (both Fs < 1), with frequencies computed from the *Frequency Dictionary of Contemporary Serbian Language* (Kostíc, 1999) and the *Corpus of Serbian Language* (Kostíc, 2001). Targets were also matched as closely as possible across the length manipulation on initial phoneme (or, in two cases, phonetic class).

Each version of the experiment contained 120 unique fillers (60 words and 60 nonwords) that could be read in only a single alphabet. These fillers comprised unique Roman items in the Roman version of the experiment and unique Cyrillic items in the Cyrillic version of the experiment. These unique fillers were included to encourage participants to read the bivalent targets in the particular alphabet to which they were assigned (see also Havelka & Rastle, 2005).

Stimulus presentation and data recording were controlled by the DMDX software (Forster & Forster, 2003) running on a Pentium III personal computer. Reading aloud responses were recorded directly to the hard drive of the PC at a sampling rate of 22 kHz, via a headset microphone fitted to each participant. Recording began on presentation of each target and continued for 2 s.

Procedure. Participants were tested individually in a quiet room. They were seated approximately 16 in. from the computer monitor and were asked to read aloud instructions from the screen before they proceeded with the experiment. Instructions were printed in the alphabet that corresponded to the (Roman or Cyrillic) version of the experiment to which participants were assigned. Participants were told that they would be seeing a series of Roman or Cyrillic (as appropriate) words and nonwords and that these should be read aloud as quickly and accurately as possible. Each trial began with a 500-ms presentation of a fixation point in the center of the screen; this was followed immediately by the presentation of the target or filler word, which stayed on screen for 2,000 ms. Participants were given 16 practice trials before the experiment began, and each participant received a different random order of the stimuli.

Results

Reaction time (RT) measurements were accomplished manually for each recorded utterance via visual inspection of the speech waveform or sound spectrogram with the Cool Edit speech-signal-processing package. The onset of acoustic energy was marked using the labeling criteria described in Rastle, Croot, Harrington, and Coltheart (2005). Acoustic labeling for 40 utterances in the data set could not be achieved because of a poor signal-to-noise ratio, and so these data points were not included in the analysis.

Data were cleaned for outliers in three ways. First, six outlying participants were excluded from each participant group, either because of nonword errors exceeding 30% or because of an average nonword RT exceeding 1,000 ms.[3] Second, one nonword item (*KAMP*) was removed from the Cyrillic list because it yielded 45% errors. Finally, three further outlying data points exceeding 1,500

[3] Six of these twelve participants were excluded on the basis of accuracy. Most of the errors of these participants (73%) involved pronouncing the ambiguous letters in nonword stimuli in the alternative script. These data suggest that despite being instructed to read using only one alphabet, these excluded participants may have had trouble inhibiting pronunciations from the other alphabet.

ms were removed. RT data for incorrect responses were also removed.

Remaining target RT and error data were analyzed by subjects and by items. The by-subjects analysis treated length (three levels) and lexical status (two levels) as repeated factors and treated alphabet in which the experiment was conducted (two levels) as an unrepeated factor. The by-items analysis treated lexical status (two levels) as a repeated factor and treated length (three levels) and alphabet in which stimuli were words (two levels) as unrepeated factors. Mean data from the by-subjects analysis are presented in Table 1.

The analyses of RT data revealed a main effect of length, $F_1(2, 76) = 77.06$, $p < .01$, $MSE = 2{,}029.67$; $F_2(2, 53) = 27.59$, $p < .01$, $MSE = 2{,}683.46$; a main effect of lexical status, $F_1(1, 38) = 162.16$, $p < .01$, $MSE = 4{,}785.30$; $F_2(1, 53) = 89.79$, $p < .01$, $MSE = 4{,}362.50$; and the predicted interaction between length and lexical status, $F_1(2, 76) = 13.16$, $p < .01$, $MSE = 2{,}497.05$; $F_2(2, 53) = 3.41$, $p < .05$, $MSE = 4{,}362.50$. This interaction indicates that the length effect on RT was substantially larger when stimuli were treated as nonwords than when they were treated as words, though the length effect nevertheless remained significant for words, $F_1(2, 76) = 15.52$, $p < .01$, $MSE = 1{,}511.78$; $F_2(2, 53) = 4.62$, $p < .05$, $MSE = 2{,}465.80$. No other effects in the analysis of RTs reached significance both by subjects and by items.

The analyses of error data revealed a main effect of length, $F_1(2, 76) = 19.16$, $p < .01$, $MSE = .004$; $F_2(2, 53) = 9.63$, $p < .01$, $MSE = 0.003$; a main effect of lexical status, $F_1(1, 38) = 46.06$, $p < .01$, $MSE = 0.005$; $F_2(1, 53) = 27.59$, $p < .01$, $MSE = 0.004$; and an interaction between length and lexical status, $F_1(2, 76) = 14.34$, $p < .01$, $MSE = 0.004$; $F_2(2, 53) = 6.65$, $p < .01$, $MSE = 0.004$. This interaction indicates that the effect of length once again was substantially larger when stimuli were treated as nonwords than when they were treated as words (though in this case the length effect did not persist for words alone ($F_1 < 1$; $F_2 < 1$). Many of the errors for nonwords (42%) involved pronouncing the ambiguous letter or letters in the alternative script. No other effects in the analysis of errors reached significance both by subjects and by items.

Discussion

In this experiment, we capitalized on the special properties of the Serbian writing system to examine the interaction between length and lexical status, in a situation in which stimuli across the manipulation of lexical status were orthographically identical. Separate groups of participants read aloud sets of Roman or Cyrillic words and nonwords. Bivalent stimuli varied on length—interpretable in one alphabet as words and in the other alphabet as nonwords—were shared across these sets. Our prediction based on models of reading aloud that postulate serial phonological assembly (e.g., Coltheart et al., 2001; Perry et al., 2007) was that the length effect would be larger when bivalent stimuli were interpreted as nonwords than when they were interpreted as words. Results were clearly in line with this prediction. For both Roman readers and Cyrillic readers, the length effect was over twice as large when the bivalent targets were treated as nonwords than when they were treated as words. Thus, it cannot be argued that the length effect is larger for nonwords than it is for words because of some visual or orthographic difference between word and nonword stimuli.

Despite the clarity of these results, one aspect of the design of this experiment was nonoptimal. Because stimuli across the manipulation of lexical status were orthographically identical, it should be possible to compare directly each individual stimulus in its word and nonword forms (i.e., compare Roman words with Cyrillic nonwords and Cyrillic words with Roman nonwords). However, though the interaction between length and lexical status reached significance in the analysis by items and, further, though this interaction was not modulated by a three-way interaction between length, lexical status, and alphabet, an inspection of the means in Table 1 shows that this kind of comparison is difficult. Specifically, it appears as if the interaction between length and lexical status held for bivalent targets that were words in Cyrillic but not for bivalent targets that were words in Roman. The problem here is that our Cyrillic participants were generally faster than the Roman participants (Roman mean RT = 717 ms; Cyrillic mean RT = 642 ms), $t(38) = 2.26$, $p < .05$, and their length effects were generally smaller than those of the Roman participants (average Roman length effect = 64 ms/letter; average Cyrillic length effect = 24 ms/letter), $t(38) = 4.76$, $p < .01$.[4] These differences between the groups could have occurred by chance or could reflect some systematic discrepancy between the two alphabets; for example, bivalent words are slightly more common in Cyrillic (2.7% of Cyrillic words) than in Roman (0.6% of Roman words). Irrespective of the reason, our second experiment avoided this difficulty by manipulating the script factor within participants.

The results of Experiment 1 rule out the possibility that the interaction between length and lexical status arises because of some visual or orthographic difference between words and nonwords. However, it remains possible that this interaction arises because of some articulatory difference between words and nonwords, which had different pronunciations of the ambiguous letters in this experiment. Experiment 2 addressed this issue by taking advantage of the special properties of Japanese kana.

Table 1
Mean Target Reaction Time (RT) and Error Data (in Parentheses) by Subjects for Experiment 1

	Roman readers		Cyrillic readers	
Condition	Words	Nonwords	Words	Nonwords
Length				
4 letters	618 (0.5%)	694 (2.0%)	580 (1.5%)	664 (4.4%)
5 letters	653 (0.5%)	767 (5.5%)	591 (2.0%)	681 (5.0%)
6 letters	687 (1.5%)	881 (16.0%)	607 (2.0%)	731 (11.0%)
Length effect	34 ms/letter	93 ms/letter	14 ms/letter	34 ms/letter

Note. Length effects (ms/letter) were calculated by computing the slope of the function relating RT to length in each condition for each participant separately and then averaging the slopes.

[4] The reason that the interaction between length and lexical status is significant in the by-items analysis and is not further modulated by alphabet is that the analysis takes into account the fact that Cyrillic readers are faster overall.

Experiment 2

Japanese kana comprises hiragana and katakana scripts, both of which are syllabic scripts characterized by perfect print-to-sound consistency at the single word level. In this writing system, it is possible to express any single spoken word in hiragana or katakana; however, in everyday writing there are a large number of spoken words that appear normally in only one of the scripts (see Wydell, Patterson, & Humphreys, 1993, for further details). This property makes it possible to present a single Japanese spoken word in a printed form that is familiar (i.e., as a word in the kana script in which it is normally written) and in a printed form of equivalent length that is not familiar (i.e., as a nonword in the kana script in which it is not usually written). Note that, unlike in Serbian, there is no alternative pronunciation to inhibit when one is given either of these printed forms; both are translated straightforwardly to the same Japanese word.

Participants in Experiment 2 were thus given two blocks of stimuli, one block of hiragana items and one block of katakana items, for reading aloud. These blocks contained (a) Japanese words appearing in their usual, orthographically familiar form (i.e., as words) and (b) Japanese words transcribed from their usual script and therefore appearing in an orthographically unfamiliar form (i.e., as nonwords). Stimuli were counterbalanced across two versions of the experiment, so no single participant produced the same spoken word twice but so exactly the same spoken words could be used across the manipulation of lexical status. Models that postulate a serial nonlexical component predict that the interaction between length and lexical status will persist under these strict conditions. Models that operate solely in parallel do not.

Method

Participants. Twenty-one Japanese native speakers were recruited from Macquarie University. All of them were educated in Japan at least to the age of 18, and the majority either held degrees from Japanese universities ($n = 12$) or were on exchange from Japanese universities ($n = 6$). Participants all had normal or corrected-to-normal vision and were free of any known reading impairments. They were paid AU$15 for their time and travel expenses.

Materials and apparatus. In total, 832 target stimuli were selected for each of four conditions: hiragana words ($n = 208$), katakana transcriptions of these hiragana words ($n = 208$), katakana words ($n = 208$), and hiragana transcriptions of these katakana words ($n = 208$). We consulted the NTT Psycholinguistic Database (Shigeki & Kondo, 1999) to determine the orthographic appropriateness of the katakana and hiragana target stimuli (e.g., to ensure that those stimuli in the word conditions would be accepted as orthographically familiar). Though the transcriptions sounded identical to the actual hiragana and katakana words that we selected, they consisted of orthographic strings that would have been unfamiliar to participants. Stimuli were divided into equal groups of 3, 4, 5, or 6 morae, with initial morae matched across length conditions.

Stimuli in each condition were divided into two equal lists for counterbalancing purposes. Each participant saw only half of the stimuli (416 items in total) to ensure that he or she never produced the same phonological string twice (i.e., to ensure that the participant did not read aloud a target word and its transcription). Participants were presented with the katakana and hiragana stimuli in separate blocks of trials, with the order of these blocks counterbalanced across participants. Stimuli were presented in separate blocks in order to minimize any potential strategic effects related to switching between scripts.

Stimulus presentation and data recording were controlled by the DMDX software (Forster & Forster, 2003) running on a Pentium III personal computer. Reading aloud responses were recorded directly to the hard drive of the PC at a sampling rate of 22 kHz, via a headset microphone fitted to each participant. Recording began on presentation of each target and continued for 3 s.

Procedure. Participants were tested individually in a quiet room. They were seated approximately 16 in. from the monitor and were told that they would be seeing a series of hiragana or katakana (as appropriate) words and nonwords that should be read aloud as quickly and accurately as possible. Participants were given a practice set prior to the presentation of each block that consisted of six hiragana or katakana stimuli. Each trial began with a fixation cross that appeared in the center of the screen for 500 ms; this fixation cross was replaced immediately by a target stimulus that was visible for 3,000 ms. Each participant received a different random order of the stimuli.

Results

RT measurements were accomplished manually for each recorded utterance via visual inspection of the speech waveform or sound spectrogram with the CheckVocal speech-signal-processing package (Protopapas, 2007). The same labeling criteria were used as in Experiment 1. The labeling of acoustic onsets for 1 participant proved impossible because of a heavy stutter, and so these data were not considered further.

Resulting data were cleaned for outliers in the same manner as in Experiment 1. The application of these criteria led to the removal of 1 participant with an average nonword RT greater than 1,000 ms. One hiragana transcription was also excluded (ま ち ね) because of an average error rate of 70%. Finally, four outlying data points over 1,500 ms were excluded from the analyses. Latencies for incorrect responses were also removed.

Remaining latency and error data were analyzed by subjects and by items using mixed-design analyses of variance that included lexical status (orthographically familiar vs. orthographically unfamiliar), length (3, 4, 5, or 6 morae), base alphabet (hiragana vs. katakana), and list (two levels) as factors. The analysis by subjects treated lexical status, length, and base alphabet as repeated factors and treated list as an unrepeated factor. The analysis by items treated lexical status as a repeated factor and treated length, base alphabet, and list as unrepeated factors. Mean data from the by-subjects analysis are presented in Table 2.

Results of the latency analyses revealed main effects of length, $F_1(3, 51) = 80.60$, $p < .01$, $MSE = 2{,}711$; $F_2(3, 399) = 59.73$, $p < .01$, $MSE = 10{,}314$, and lexical status, $F_1(1, 17) = 131.64$, $p < .01$, $MSE = 2{,}492$; $F_2(1, 399) = 398.60$, $p < .01$, $MSE = 2{,}278$, and, critically, the predicted interaction between these two variables, $F_1(3, 51) = 35.01$, $p < .01$, $MSE = 471$; $F_2(3, 399) = 20.13$, $p < .01$, $MSE = 2{,}278$. This interaction reflects the fact that the length effect was substantially larger in the context in which kana stimuli were transcribed into an orthographically unfamiliar

Table 2
Mean Target Reaction Time (RT) and Error Data (in Parentheses) by Subjects for Experiment 2

	Katakana stimuli		Hiragana stimuli	
Condition	Words	Nonwords	Words	Nonwords
Length				
3 morae	590 (1.8%)	631 (3.6%)	591 (3.1%)	615 (3.4%)
4 morae	607 (1.5%)	671 (1.5%)	607 (1.6%)	643 (1.6%)
5 morae	639 (1.7%)	756 (5.2%)	648 (1.1%)	703 (1.5%)
6 morae	671 (1.7%)	797 (8.1%)	686 (2.8%)	750 (3.3%)
Length effect	27 ms/mora	59 ms/mora	32 ms/mora	46 ms/mora

Note. Length effects (ms/mora) were calculated by computing the slope of the function relating RT to length in each condition for each participant separately and then averaging the slopes. Nonwords in each script condition are transcriptions of stimuli that are words in the alternative script.

form than in the context in which they were presented in an orthographically familiar form. Nevertheless, the length effect persisted for word stimuli alone, $F_1(3, 51) = 44.08$, $p < .01$, $MSE = 1{,}326$; $F_2(3, 400) = 27.15$, $p < .01$, $MSE = 6{,}011$. There were no other interactions involving length that reached significance both by subjects and by items.

The percentage of incorrect responses in this experiment was very low ($M = 2.80\%$; see Table 2). Nevertheless, the error analyses revealed main effects of length, $F_1(3, 51) = 4.08$, $p < .05$, $MSE = 0.002$; $F_2(3, 399) = 2.98$, $p < .05$, $MSE = 0.006$, and lexical status, $F_1(1, 17) = 8.84$, $p < .01$, $MSE = 0.002$; $F_2(1, 399) = 25.71$, $p < .01$, $MSE = 0.003$. However, the interaction between these latter two factors failed to reach significance by items, $F_1(3, 51) = 3.89$, $p < .05$, $MSE = 0.001$; $F_2(3, 399) = 1.96$, $p > .10$, $MSE = 0.003$. The only indication of an interaction between length and lexical status in the error data was for the katakana stimuli, as revealed by a three-way interaction between length, lexical status, and script, $F_1(3, 51) = 3.26$, $p < .05$, $MSE = 0.001$; $F_2(3, 399) = 4.95$, $p < .01$, $MSE = 0.003$.

Discussion

In this experiment, we took advantage of the special properties of the Japanese writing system to investigate the interaction between length and lexical status, in a situation in which stimuli across the manipulation of lexical status were phonologically identical. Japanese speakers read aloud separate sets of hiragana and katakana stimuli that varied in length. Stimuli in each of these sets comprised (a) words printed in the script in which they normally appear and (b) nonwords normally printed in the other script. Our prediction, based on models of reading aloud that postulate a serial nonlexical component, was that the length effect would be larger when the Japanese words were printed in a script in which they do not usually appear than when they were printed in the script in which they normally appear. Results revealed clearly that the length effect was substantially larger when the Japanese targets were printed in an orthographically unfamiliar form than when they were printed in an orthographically familiar form. These data thus rule out the argument that the length effect is larger for nonwords than it is for words because of some phonological or articulatory difference between word and nonword stimuli.

General Discussion

The question of whether the transformation from orthography to phonology involves serial processing has been one of especially vigorous debate over the past 10 years, as some researchers have argued strongly that the assembly of phonological information occurs in a left-to-right manner (e.g., Coltheart & Rastle, 1994; Coltheart et al., 2001; Rastle & Coltheart, 1999; Roberts et al., 2003; see Rastle & Coltheart, 2006 for a review) and others have argued equally strongly that this transformation arises solely in parallel (e.g., Plaut et al., 1996; Seidenberg & McClelland, 1989; Zorzi, 2000). Numerous data sets point to the existence of a serial process in reading (e.g., Havelka & Rastle, 2005; Rastle & Coltheart, 1998, 1999; Weekes, 1997). However, this issue has been difficult to resolve because (a) it has sometimes been shown that these serial effects can be simulated by models that operate solely in parallel, thus implying some confound between the serial manipulation of interest and another variable to which parallel models are sensitive (see, e.g., Plaut et al., 1996; Zorzi, 2000), and (b) it has been argued that serial effects arise in peripheral aspects of the reading system (e.g., visual analysis, articulatory processing) that everyone agrees have a serial component.

The experiments reported here advance this debate by investigating the robustness of the interaction between length and lexical status (Weekes, 1997). Though this effect has been used in support of models of reading aloud that incorporate a serial component (e.g., Coltheart et al., 2001), the interpretation of this effect is complicated by the fact that there are numerous potential orthographic and phonological differences between words and nonwords that could modulate the length effect (see Seidenberg et al., 1996, for further discussion). Our research gets around this problem by investigating the interaction between length and lexical status in situations in which identical orthographic (Experiment 1) and phonological (Experiment 2) strings can be used across the manipulation of lexical status. In Experiment 1 we investigated the length effect in Serbian words and nonwords, using the same bivalent targets in both of these contexts. Similarly, in Experiment 2 we investigated the length effect in Japanese words and nonwords, using the same spoken words in both of these contexts. Results were unambiguous. They revealed that the length effect was around twice as large when stimuli were presented in a nonword context than when they were presented in a word context. These data allow us to rule out the possibility that differences in the length effect across the manipulation of lexical status arise in peripheral aspects of the reading system concerned with visual or articulatory processing (Seidenberg et al., 1996; Seidenberg & Plaut, 1998).

One aspect of our data may seem inconsistent with previous reports. In particular, in both of our studies we observed a strong length effect for words alone, whereas Weekes (1997) reported a null effect of length for these items (see also Bijeljac-Babic, Millogo, Farioli, & Grainger, 2004). Further, though a number of studies have reported inhibitory effects of length for words (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Spieler & Balota, 1997; Ziegler et al., 2001; see New, Ferrand, Pallier, & Brysbaert, 2006, for a review), the effects seen in our studies were much larger than would be expected from these reports. However, it is important to recognize that our manipulation of length differs from those in these previous studies in at least two important ways. One is that these previous studies focused on monosyllabic reading, whereas all of our stimuli were polysyllabic (and, more important, had an increasing number of syllables or an increasing number of morae with increasing length). Though the topic is still a matter of some debate (Bachoud-Levi, Dupoux, Cohen, & Mehler, 1998) it does seem fairly clear that increasing the number of syllables in an utterance increases production latency (Sternberg, Monsell, Knoll, & Wright, 1978). The other difference between our work and this previous research is that we investigated writing systems characterized by perfect spelling-to-sound consistency, whereas most of this previous research was based on English. Some research suggests that nonlexical processing may be particularly strong in languages with shallower orthographies (Frost, Katz, & Bentin, 1987; Wydell, Vuorinen, Helenius, & Salmelin, 2003). If the length effect arises because of a serial nonlexical process, this too could explain why we observed a particularly strong length effect for words.

Overall, then, our findings provide some of the most powerful evidence to date for dual-route models of reading aloud in which phonological assembly is accomplished in a serial manner, such as the DRC model (Coltheart et al., 2001) and the CDP+ model (Perry et al., 2007). These findings contribute to a growing body of literature already suggestive of a serial process in the orthography-to-phonology transformation: the position of irregularity effect (Coltheart & Rastle, 1994; Rastle & Coltheart, 1999; Roberts et al., 2003), the position of bivalence effect (Havelka & Rastle, 2005), the whammy effect (Rastle & Coltheart, 1998), and the position-sensitive Stroop effect (Coltheart et al., 1999). On the other hand, our findings would seem to pose a significant challenge to models of reading that operate solely in parallel. Indeed, these findings would seem to call into question the whole PDP approach to reading aloud, given that the PDP approach is based on the a priori assumption that all cognitive processing must be parallel processing (e.g., Harm & Seidenberg, 2004; Plaut et al., 1996).

References

Bachoud-Lévi, A. C., Dupoux, E., Cohen, L., & Mehler, J. (1998). Where is the length effect? A cross-linguistic study of speech production. *Journal of Memory and Language, 39,* 331–346.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133,* 283–316.

Bijeljac-Babic, R., Millogo, V., Farioli, F., & Grainger, J. (2004). A developmental investigation of word length effects in reading using a new on-line word identification paradigm. *Reading & Writing, 17,* 411–431.

Brysbaert, M. (2003). Bilingual visual word recognition: Evidence from masked phonological priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: The state of the art* (pp. 323–343). New York: Psychology 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., Langdon, R., & Ziegler, J. (2001). The DRC model: A model of visual word recognition and reading aloud. *Psychological Review, 108,* 204–256.

Coltheart, M., Woollams, A., Kinoshita, S., & Perry, C. (1999). A position-sensitive Stroop effect: Further evidence for a left-to-right component in print-to-speech conversion. *Psychonomic Bulletin & Review, 6,* 456–463.

Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5,* 175–197.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers, 35,* 116–124.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographic depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 104–115.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Division of labor between visual and phonological processes. *Psychological Review, 111,* 662–720.

Havelka, J., & Rastle, K. (2005). The assembly of phonology from print is serial and subject to strategic control: Evidence from Serbian. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 148–158.

Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language, 44,* 2–31.

Kawamoto, A. H., Kello, C., 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, 25,* 362–381.

Kostíc, D. J. (2001). *Quantitative description of Serbian language structure: Corpus of Serbian language.* Belgrade, Serbia: Institute for Experimental Phonetics and Speech Pathology, Belgrade and Laboratory for Experimental Psychology, University of Belgrade.

Kostíc, D. J. (1999). *Frequency dictionary of contemporary Serbian language.* Belgrade, Serbia: Institute for Experimental Phonetics and Speech Pathology and Laboratory for Experimental Psychology, University of Belgrade.

Lukatela, G., Lukatela, K., Carello, C., & Turvey, M. T. (1999). Effects of frequency and phonological ambiguity on naming Serbo-Croatian words. *European Journal of Cognitive Psychology, 11,* 1–16.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review, 13,* 45–52.

Perry, C., & Ziegler, J. C. (2002). Cross-language computational investigation of the length effect in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 990–1001.

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. (1999). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science, 23,* 543–568.

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. (1998). Whammies and double whammies: The effect of length on nonword reading. *Psychonomic Bulletin & Review, 5,* 277–282.

Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 482–503.

Rastle, K., & Coltheart, M. (2006). Is there serial processing in the reading system; and are there local representations? In Andrews, S. (Ed.), *From inkmarks to ideas: Progress in visual word recognition.* Hove, England: Psychology Press.

Rastle, K., Croot, K., Harrington, J., & Coltheart, M. (2005). Characterizing the motor execution stage of speech production: Consonantal influences on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1083–1095.

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.

Roberts, M., 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.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Seidenberg, M. S., Petersen, M., MacDonald, M. C., & Plaut, D. C. (1996). Pseudohomophone effects and models of word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 48–62.

Seidenberg, M. S., & Plaut, D. C. (1998). Evaluating word reading models at the item level: Matching the grain of theory and data. *Psychological Science, 9,* 234–237.

Shigeki, S., & Kondo, K. (1999). *The NTT Japanese Psycholinguistic Database.* Tokyo: Sanseido.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8,* 411–416.

Sternberg, S., Monsell, S., Knoll, R., & Wright, C. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117–152). New York: Academic Press.

van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 616–627.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology, 50*A, 439–456.

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,* 221–243.

Wydell, T. N., Patterson, K. E., & Humphreys, G. W. (1993). Phonologically mediated access to meaning for kanji: Is a *rows* still a *rose* in Japanese kanji? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 491–514.

Wydell, T. N., Vuorinen, T., Helenius, P., & Salmelin, R. (2003). Neural correlates of letter-string length and lexicality during reading in a regular orthography. *Journal of Cognitive Neuroscience, 17,* 1052–1062.

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science, 12,* 379–384.

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.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1131–1161.

Received October 5, 2007
Revision received July 25, 2008
Accepted September 17, 2008 ■