Memory & Cognition
2005, 33 (1), 76-85

Graphemic complexity and multiple print-to-sound associations in visual word recognition

ARNAUD REY
LEAD–CNRS, Université de Bourgogne, Dijon, France

and

NIELS O. SCHILLER
Maastricht University, Maastricht, The Netherlands
and Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

It has recently been reported that words containing a multiletter grapheme are processed slower than are words composed of single-letter graphemes (Rastle & Coltheart, 1998; Rey, Jacobs, Schmidt-Weigand, & Ziegler, 1998). In the present study, using a perceptual identification task, we found in Experiment 1 that this graphemic complexity effect can be observed while controlling for multiple print-to-sound associations, indexed by regularity or consistency. In Experiment 2, we obtained cumulative effects of graphemic complexity and regularity. These effects were replicated in Experiment 3 in a naming task. Overall, these results indicate that graphemic complexity and multiple print-to-sound associations effects are independent and should be accounted for in different ways by models of written word processing.

The nature and structure of print-to-sound conversion processes is one of the most debated issues in visual word recognition. Several computational models of word reading have been proposed and provide detailed descriptions of these processes (e.g., Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Jacobs, Rey, Ziegler, & Grainger, 1998; Norris, 1994; Plaut, McClelland, Seidenberg, & Patterson, 1996; Zorzi, Houghton, & Butterworth, 1998). In alphabetic writing systems, this issue is particularly interesting and challenging because of several print-to-sound problems that readers learn to solve and master during reading acquisition. In the present study, we will particularly focus on two major print-to-sound problems: the graphemic complexity problem and the multiple print-to-sound associations problem. Our purpose is to demonstrate that these problems reflect distinct processing difficulties that take place at different levels within the reading processes involved in print-to-sound conversion.

The graphemic complexity problem results from the fact that in most alphabetic languages, the number of elementary visual units composing a word (i.e., letters) can be and often is different from the number of elementary phonological units composing the phonological form of this word (i.e., phonemes). For example, the written word BREAD is composed of five letters but its pronunciation has only four phonemes (i.e., /bred/). On the other hand, the word GRASP also has five letters and exactly the same number of phonemes (i.e., /grasp/). The challenge is to understand how readers handle such a mismatch between number of letters and number of phonemes.

The notion grapheme provides a solution to this first problem of print-to-sound conversion. *Graphemes* are defined as the orthographic correspondent of phonemes. The word BREAD, for example, has four graphemes, B, R, EA, and D, which correspond to the four phonemes /b/, /r/, /e/, and /d/. Similarly, the word GRASP has five graphemes, G, R, A, S, and P, which correspond to the five phonemes /g/, /r/, /a/, /s/, and /p/. During reading acquisition, readers learn to associate single-letter graphemes like R or A to their corresponding phonemes /r/ and /a/, and they also learn to associate multiletter graphemes like EA to the phoneme /e/. Learning these associations thus provides a solution to the mismatch between number of letters and number of phonemes.

However, establishing these various types of associations (i.e., single- and multiletter grapheme-to-phoneme associations) might generate new difficulties. Indeed, while children learn that the single-letter grapheme A is pronounced /a/ in GRASP, at the same time or soon after, they learn that the multiletter grapheme EA, which is composed of two single-letter graphemes E and A, is pronounced /e/ in BREAD. It seems, therefore, that although the notion *grapheme* solves the letters-to-phonemes mismatch, it generates a graphemic complexity problem because multiletter graphemes (like EA) are always com-

---

This research was supported by a grant from the Fyssen Foundation to A.R. N.O.S. was supported by the Royal Netherlands Academy of Arts and Sciences (KNAW) and by a "Vernieuwingsimpuls" grant from the Netherlands Organization for Scientific Research (NWO). We thank Max Coltheart for providing us with the set of rules generated by the dual-route model, Alfonso Caramazza for his support, and Virginia Savova for her help in running the experiments. Correspondence concerning this article should be addressed to A. Rey, LEAD–CNRS, Université de Bourgogne, Pôle AAFE-Esplanade Erasme, BP 26513 21065, Dijon Cedex, France (e-mail: arnaud.rey@u-bourgogne.fr).

posed of single-letter graphemes (here, E and A). Indeed, while readers have to process EA as a unit, they also have to inhibit the potential simultaneous activation of the single-letter graphemes E and A (Lange, 2002; Peereman, Brand, & Rey, in press; Rey, Ziegler, & Jacobs, 2000). Processing a letter sequence that incorporates a multiletter grapheme might, therefore, provoke either a competition between multi- and single-letter graphemes or competition between phonemes that have been activated by these different graphemes.

Due to this competition between different levels of processing units, one should observe longer processing times for words containing multiletter graphemes, compared with words composed essentially of single-letter graphemes. This result has, indeed, been described recently with skilled English adult readers in a nonword naming experiment (Rastle & Coltheart, 1998) and in a perceptual identification task done with English and French words (Rey, Jacobs, Schmidt-Weigand, & Ziegler, 1998). Although the effect has not been interpreted in the same way in these two studies, both indicate that the presence of multiletter graphemes produces a processing cost relative to single-letter graphemes in the reading system, even for skilled readers.

The second obstacle in reading mentioned above is the multiple print-to-sound associations problem. In alphabetic languages like English, there are many cases in which the same sequence of letters is pronounced in different ways. For example, the grapheme EA is pronounced /i/ in BEACH, whereas it is pronounced /e/ in BREAD. The same printed unit is in this case associated with different phonological units, and these multiple associations might generate another kind of processing competition.

This problem has, however, been addressed in different ways in the experimental literature (e.g., Glushko, 1979; Jared, 1997; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Taraban & McClelland, 1987). The differences between these approaches result from the kind of linguistic units that have been considered and the theoretical frameworks chosen by these researchers. Glushko proposed to study the effect of multiple print-to-sound associations for a subsyllabic linguistic unit called the *body*, which is composed of the vowel and the coda consonants of a syllable (e.g., ASP in GRASP or EAD in BREAD). He found that low-frequency words with inconsistent bodies (i.e., having multiple pronunciations) are processed less rapidly than are consistent words. More recently, Jared (1997) revised this notion by demonstrating that the critical factor for observing consistency effects is to consider the difference between the frequency of *friends* (other words that have the target's orthographic body and the same pronunciation of it) and the frequency of *enemies* (other words that have the target's orthographic body but a different pronunciation of it). She found that words were processed less rapidly if their body had high-frequency enemies and low-frequency friends. This result indicates that the problem of multiple print-to-sound associations is modulated by the differential strength of these associations (see also Jared, 2002).

Alternatively, other studies considered the multiple-associations problem at the level of graphemes and introduced the notion of *regularity* (e.g., Coltheart, 1978; Coltheart et al., 1993; Cooltheart et al., 2001). The most frequent association between a grapheme and a phoneme is defined in these studies as the *regular* pronunciation of this grapheme. In several experiments, it was reported that low-frequency irregular words (i.e., words containing a grapheme with a rare pronunciation) were processed less rapidly than were regular words (e.g., Seidenberg et al., 1984; Taraban & McClelland, 1987). In a sense, the definition of regularity is similar to the definition of *consistency* given by Jared (1997). Multiple associations cause a processing problem when a word's unit has one higher frequency enemy (for graphemic regularity) or has a set of high-frequency enemies (for body consistency). In both cases, apart from the choice of orthographic unit, the crucial factor is the existence of at least one stronger association from orthography to phonology. Whether the consistency of the body or the regularity of graphemes produces processing problems for the reading system will not be directly debated in this article (see Coltheart et al., 2001, and Plaut et al., 1996, for opposite views on this issue). Here, we decided to restrict our analysis of the multiple-associations problem to the level of graphemes and to maintain constant the body consistency of our test items.

Together, these two print-to-sound problems—namely, the graphemic complexity problem and the multiple print-to-sound associations problems—provide important constraints for modeling word recognition. There remain, however, two related issues that need to be clarified. First, it might be possible that the multiletter grapheme effect that has been reported recently (Rastle & Coltheart, 1998; Rey et al., 1998) reflects in fact a multiple print-to-sound associations problem. Indeed, words with multiletter graphemes like BREAD might be responded to less rapidly than words composed of single-letter graphemes like GRASP, because multiletter graphemes like EA are associated to several phonemes. Since no index of regularity or consistency was controlled in experiments showing a multiletter grapheme effect for words, this effect could therefore be interpreted as a multiple print-to-sound association's effect. Experiment 1 was conducted to verify whether a multiletter grapheme effect remains when words are controlled for graphemic regularity and body consistency.

The second question concerns the relation between the multiletter grapheme effect and the multiple-associations effect. In principle, these effects should be independent because they are theoretically due to different processing difficulties. The multiletter grapheme effect results from competition between multiletter graphemes and single-letter graphemes, whereas the multiple-associations effect is related to the wrong activation of a phoneme that is more frequently associated with a given grapheme. One should therefore observe cumulative effects of graphemic complexity and multiple associations: Processing times should be longer for irregular words with multiletter graphemes, as compared with regular words composed of single-letter graphemes. This prediction was tested in Experiment 2 with a perceptual identification task and in Experiment 3 with a naming task.

For testing the hypotheses of Experiments 1 and 2, we used the same experimental paradigm as in Rey et al. (1998)—namely, a perceptual identification task. In this task, participants are presented with a black computer screen from which words emerge progressively. This progressive increase in a word's visibility is done by changing the word's color from black to various types of gray and finally to white. Participants simply press the space bar of the keyboard as soon as they have identified the word. They are then asked to enter the identified word on the keyboard. The elapsed time between the beginning of the increase in a word's visibility and the pressing of the space bar is the dependent variable and is considered a word's *identification time*. An interesting feature of this experimental paradigm therefore is that no phonological output is produced contrary to the standard naming task. This has major consequences for testing computational models of visual word recognition. Indeed, if the graphemic complexity and the multiple-associations effects are observed in this paradigm, any interpretation of these effects in terms of phonological output appears to be inadequate in the present situation.

EXPERIMENT 1

Method

Participants. Twenty-three undergraduate students at Harvard University participated in the experiment. All were native English speakers and had normal or corrected-to-normal vision.

Stimuli and Apparatus. Two groups of 15 monosyllabic English five-letter words were selected (see Appendix A). One group contained words composed of three phonemes (e.g., TEETH → /tiθ/) and the other group contained words composed of five phonemes (e.g., CRISP → /krɪsp/). Frequency was estimated using the CELEX frequency count (Baayen, Piepenbrock, & Gulikers, 1995). The two groups were matched for word frequency (7.1 vs. 7.2 occurrences per million for three- and five-phoneme words, respectively), summed bigram frequency (4,812 vs. 5,139 occurrences per million), orthographic neighborhood density (2.13 vs. 2.53), number of higher frequency neighbors (1.47 vs. 1.67), and phonological neighborhood density (1.93 vs. 1.6). *t* tests were systematically conducted in order to verify that the two word groups did not differ on the preceding set of variables and that none of these tests reached significance (all ts < 1). All words were regular according to Coltheart's (1978) set of rules and were feedforward consistent, their orthographic rimes being always pronounced in the same way.

The experiment was controlled by an IBM PC 486 DX2 computer. The stimulus words were typed in lowercase. The experiment was run in a darkened room that was lit with a lamp placed behind the participants. The contrast of the screen was set at its maximum; that is, the background was as dark as possible. Stimulus luminance, on the other hand, was set to be as high as possible.

Procedure. Each trial began with a 1-sec presentation of a fixation mark ("+") in the center of the screen. The fixation mark was replaced by the target word, which was written in black (i.e., completely invisible, the background also being black). The luminance of the target word was then progressively increased by modifying the color of the target word. This was done by incrementing every 100 msec the values of the RGB (red, green, blue) counters of one unit. Thus, every counter was set at 0 at the beginning. After 100 msec, the red counter was set at 1 (the green and blue counters still being at 0). After 200 msec, the RGB counters were at 1–1–0, respectively. After 300 msec, the RGB counters were at 1–1–1, after 400 msec, the RGB counters were at 2–1–1, and so forth. As soon as the participants could identify the target word, they interrupted the luminance-increasing process by pressing the space bar. Then, the item was replaced by a pattern mask and participants were asked to enter what they had seen by using the keyboard. After this, they pressed the "return" key, and the screen remained black for 500 msec until the next trial started. For each trial, response time was recorded (that is, the time interval between the onset of the luminance-increasing procedure and the pressing of the space bar). Participants were instructed to stress accuracy rather than speed.

Results

Mean correct response times and error rates for the three experimental conditions are reported in Table 1. The trimming procedure excluded six data points greater than three *SD*s above and below the participants' overall mean response time (four were from the three-phoneme condition and two from the five-phoneme condition). Analyses of variance (ANOVAs) were conducted, with target type (three or five phonemes) as the independent within-participants variable, using both participants (F_1) and items (F_2) as random factors.

For response times, the main effect of target type was significant [$F_1(1,22) = 11.03, p < .01; F_2(1,28) = 9.72, p < .01$]. Response times were 49 msec faster for five-phoneme words (2,900 msec), compared with three-phoneme words (2,949 msec). For errors, there was no difference between the two experimental conditions (both Fs < 1).

In order to strengthen the present result, we conducted a multiple-regression analysis, entering item latencies as a dependent variable and the number of phonemes (i.e., three or five phonemes, the two categories of our experimental manipulation), word frequency, summed bigram frequency, orthographic neighborhood density, number of higher frequency neighbors, and phonological neighborhood density, as independent variables. Altogether, these variables explained 32% of the variance [$F(6,23) = 1.8, p = .14$]. More importantly, only the number of phonemes explained a unique and significant part of the variance [$t(23) = 2.05, p = .05$], while other variables did not (all ts < 1).

Discussion

Experiment 1 was designed to verify whether the multi-letter grapheme effect observed by Rastle and Coltheart (1998) and Rey et al. (1998) could in fact be interpreted

Table 1
Mean Response Times (RTs, in Milliseconds), Percentages of Error, and Standard Deviations (*SD*s) for Words Having Three or Five Phonemes in Experiment 1

	RTs		% Error	
Words	*M*	*SD*	*M*	*SD*
Three phonemes	2,949	45	0.87	0.48
Five phonemes	2,900	39	1.16	0.54

as a multiple print-to-sound associations effect. For this purpose, the same manipulation as in Rey et al. (1998) was conducted on the number of phonemes with a set of regular and consistent words. The data indicate that the multiletter grapheme effect can be observed independently of the multiple-associations effect. This result provides additional evidence for the competition between various levels of graphemic units during visual word processing.

Although Experiment 1 provides a clear argument favoring the independence of the multiletter grapheme effect and the multiple-associations effect, another fundamental empirical argument would be to observe these effects simultaneously within the same experiment. Experiment 2 was therefore designed to test the hypothesis of a cumulative effect of both factors. The multiple-associations factor was addressed in this experiment by using the regularity index. Given that regular words are composed of the most frequent grapheme-to-phoneme associations and irregular words are composed of at least one low-frequency grapheme-to-phoneme association, we selected these categories in order to have two groups of words varying in their level of print-to-sound associations. Unfortunately, due to constraints in the linguistic material, it was impossible to manipulate factorially both graphemic complexity and regularity. We therefore decided to compare perceptual identification performances on three groups of words: three-phoneme regular words, three-phoneme irregular words, and five-phoneme regular words (given that there were too few five-phoneme irregular words). Note that these three groups were all feedforward consistent (i.e., their body is always pronounced in the same way).

EXPERIMENT 2

Method

Participants. Twenty-seven undergraduates at Harvard University participated in the experiment. All were native English speakers and had normal or corrected-to-normal vision.

Stimuli and Apparatus. Three groups of 14 monosyllabic English five-letter words were selected. A first group contained regular words composed of three phonemes (e.g., SHOAL → /ʃƈəl/; diphthongs and affricates are counted as one phoneme; for phonetic evidence, see Kenstowicz, 1994, p. 46), a second group contained irregular words composed of three phonemes (e.g., ROUTE → /rut/), and a third group contained regular words composed of five phonemes (e.g., CRISP → /krɪsp/). Frequency was estimated using the CELEX frequency count (Baayen et al., 1995). The three groups were matched as closely as possible for word frequency, summed bigram frequency, orthographic neighborhood density, number of higher frequency neighbors, and phonological neighborhood density (see Appendix B for a detailed description of these statistics). t tests comparing each group with the other two groups on these psycholinguistic dimensions did not reveal any significant difference (all ts < 1). Regularity was defined according to Coltheart's set of rules (Coltheart et al., 1993). The experiment was controlled by an IBM PC 486 DX2 computer. The experimental setup and procedure were identical to that used in Experiment 1.

Results

Mean correct response times and error rates for the three experimental conditions are reported in Table 2. The trimming procedure excluded 6 data points greater than three *SD*s above and below the participants' overall mean response time. These data points were evenly distributed among the three experimental conditions (i.e., 2–1–3). ANOVAs were conducted, with target type (three-phoneme regular, three-phoneme irregular, and five-phoneme regular) as the independent within-participants variable, using both participants (F_1) and items (F_2) as random factors.

For response times, we observed a main effect of target type [$F_1(2,52) = 5.17, p < .01$; $F_2(2,39) = 3.76, p < .05$]. Response times were faster for five-phoneme regular words (2,872 msec), compared with three-phoneme regular words (2,908 msec) and three-phoneme irregular words (2,936 msec). Planned comparisons indicated that there was a significant difference between three- and five-phoneme words [$F_1(1,26) = 8.41, p < .01$; $F_2(2,39) = 6.08, p < .05$], five-phoneme words being responded to faster than were three-phoneme words. On the other hand, there was a significant difference between irregular words and regular words [$F_1(1,26) = 7.07, p = .01$; $F_2(2,39) = 5.18, p < .05$], regular words being responded to faster than were irregular words. For errors, there was no difference between the three experimental conditions (both Fs < 1).

As in Experiment 1, two multiple-regression analyses were conducted in order to verify that both the graphemic complexity and the regularity effects could not be accounted for by potentially confounded variables. For this purpose, item latencies were entered in the regressions as a dependent variable, and the same set of variables as in Experiment 1 were entered as independent variables (i.e., word frequency, summed bigram frequency, orthographic neighborhood density, number of higher frequency neighbors, and phonological neighborhood density). First, when the number of phonemes was included in the set of independent variables, these variables explained altogether 22.8% of the variance [$F(6,35) = 1.73, p = .14$], but only the number of phonemes accounted for a unique and significant part of the variance [$t(35) = 2.42, p < .05$]. Second, when a binary variable coding for regularity was included in the set of independent variables, all these variables explained 22.7% of the variance [$F(6,35) = 1.72, p = .15$], but only regularity accounted for a unique and significant part of the variance [$t(35) = 2.4, p < .05$]. These analyses therefore rule out alternative accounts of the present effects in terms of confounded variables (at least for the set of variables that has been considered here).

Table 2
Mean Response Times (RTs, in Milliseconds), Percentages of Error, and Standard Deviations (*SD*s) for Three-Phoneme Irregular Words, Three-Phoneme Regular Words, and Five-Phoneme Regular Words in Experiment 2

	RTs		% Error	
Words	*M*	*SD*	*M*	*SD*
Three-phoneme irregular	2,936	42	2.9	0.87
Three-phoneme regular	2,908	30	2.4	0.76
Five-phoneme regular	2,872	34	1.9	0.72

Discussion

While Experiment 1 demonstrated that the graphemic complexity effect is present when the regularity or the consistency of print-to-sound associations are controlled, Experiment 2 further showed that effects of graphemic complexity and multiple print-to-sound associations can be observed jointly in the same experiment. This indicates that words with multiletter graphemes take longer to identify, as compared with words essentially composed of single-letter graphemes, and that additional processing time is required if one of a word's graphemes is not pronounced with the most frequent grapheme-to-phoneme association.

Apart from the cumulative effect of graphemic complexity and multiple print-to-sound associations, it must be noted that the present results have been obtained in a perceptual identification task that does not require any overt pronunciations from participants. Although phonological representations are likely activated in this paradigm, participants do not have to overtly pronounce the target words. Therefore, since reading aloud was not required, it is possible to rule out any explanation of these effects at the level of phonological output.

The main difference between a perceptual identification task and a naming task comes from response procedures. Whereas the present perceptual identification task requires a buttonpress and the same motor response on each trial, reading aloud requires a different output response for each word and a different sequence of articulatory motor programs. Presumably, graphemic complexity and multiple print-to-sound effects result from processing difficulties occurring before the phonological output level. Therefore, these effects should also be present in a naming task. This prediction is tested in Experiment 3.

EXPERIMENT 3

The same stimuli as in Experiment 2 were used in Experiment 3. Since the present set of stimuli was not matched for initial phoneme, a factor that is now known to affect naming latencies (e.g., Kessler, Treiman, & Mullennix, 2002; Spieler & Balota, 1997; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995), participants systematically performed a delayed naming task on the same stimuli after the naming experiment. Delayed naming is supposed to capture variance related to factors involved in item pronunciation. Latencies obtained in delayed naming can therefore be used as a covariate in the analysis of naming latencies in order to control variance related to output pronunciation.

Method

Participants. Twenty-nine undergraduate students at Harvard University participated in the experiment. All were native English speakers and had normal or corrected-to-normal vision.

Stimuli and Apparatus. The same stimuli as in Experiment 2 were used. The experiment was controlled with the PsyScope program (Cohen, MacWhinney, Flatt, & Provost, 1993) on a Power Macintosh computer.

Procedure. The participants systematically performed a naming task followed by a delayed naming task. The same stimuli were used in both tasks and were presented in a different random order for each participant.

In the naming task, a trial started by the presentation of a fixation point (":") for 700 msec. It was followed by a blank screen for 500 msec, and by a target word that remained visible until the participant's response. Words were displayed in the middle of the screen in 36-point Geneva bold font. Response times were recorded from the onset of target presentation to the trigger of a voice key. This sequence was followed by an intertrial interval (ITI) of 1 sec. The experimenter registered erroneous pronunciations, but no feedback was provided during the experiment.

In the delayed naming task, a trial started by the presentation of a fixation point (":") for 700 msec. It was followed by a blank screen for 500 msec and by a target word that remained visible for 1,500 msec. The same font and size as in the naming task was used to display target words. The word was followed by a blank screen for 200 msec and by a go-cue ("*") that remained on the screen until the participant's response. The participants were instructed to pronounce the target word as soon as the cue appeared on the screen. Response times were recorded from the onset of the cue to the trigger of a voice key. This sequence was followed by an ITI of 1 sec. Erroneous pronunciations were registered, but no feedback was provided.

Results

Mean correct response times and error rates for the three experimental conditions in the naming and delayed naming tasks are reported in Table 3. In the naming task, the trimming procedure excluded 13 data points greater than three *SD*s above and below the participants' overall mean response time. There were 10 outliers in the three-phoneme irregular condition, 1 in the three-phoneme regular condition, and 2 in the five-phoneme regular condition. In the delayed naming task, before applying the trimming procedure, 2 participants were excluded due to a high rate of outliers (45% of RTs higher than 1,000 msec) for one of them and a high rate of anticipations (48% of RTs smaller than 150 msec) for the other. Because no score was available in delayed naming for these participants, they were also excluded from the naming analyses. For the remaining 27 participants, the same trimming procedure was applied to the delayed naming data as for naming data. Four outliers were evenly distributed across conditions (i.e., 2–1–1). ANOVAs were conducted with target type (same conditions as in the previous experiment) as the independent within-participants

Table 3
Mean Response Times (RTs, in Milliseconds), Percentages of Error, and Standard Deviations (*SD*s) in Naming and Delayed Naming for Three-Phoneme Irregular Words, Three-Phoneme Regular Words, and Five-Phoneme Regular Words in Experiment 3

	Naming				Delayed Naming			
	RTs		% Error		RTs		% Error	
Words	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Three-phoneme irregular	530	13	1.85	0.7	441	22	0.7	0.4
Three-phoneme regular	511	12	1.85	0.7	444	19	0.7	0.4
Five-phoneme regular	502	10	1.60	0.7	452	25	1.2	0.5

variable, using both participants (F_1) and items (F_2) as random factors.

In the naming task, a main effect of target type on response times was only significant for participants [$F_1(2,52) = 12.15, p < .001$; $F_2(2,39) = 1.86, p = .17$]. Response times were faster for five-phoneme regular words (502 msec), compared with three-phoneme regular words (511 msec) and with three-phoneme irregular words (532 msec). Planned comparisons indicated a significant difference between three- and five-phoneme words in the participant analysis only [$F_1(1,26) = 10.66$, $p < .01$; $F_2(1,39) = 1.82, p = .19$], five-phoneme words being responded to faster than three-phoneme words. Similarly, there was a significant difference between irregular words and regular words in the participant analysis only [$F_1(1,26) = 18.13, p < .001$; $F_2(1,39) = 3.05$, $p = .07$], regular words being responded to faster than irregular words. For errors, there was no difference between the three experimental conditions (both Fs < 1). In the delayed naming task, no difference between the three experimental conditions was observed on response times or errors (all Fs < 1).

A second ANOVA was conducted on item naming latencies, using delayed naming latencies as a covariate factor. The nonsignificant effects obtained in the previous item analyses might indeed be due to an undesirable source of variance—namely, variance related to pronunciation and articulatory factors. However, this variance is a priori captured by delayed naming latencies, and these data can therefore be used as a covariate factor in order to control for this source of variance. We included delayed naming latencies as a covariate, obtaining a main effect of target type on response times that was, in this case, significant by items [$F_2(2,38) = 3.49, p < .05$]. Similarly, planned comparisons indicated that the difference between three- and five-phoneme words was marginally significant by items [$F_2(1,38) = 4.19, p = .05$], and that the difference between irregular words and regular words was significant by items [$F_2(1,38) = 6.14$, $p < .05$].

Finally, as in Experiments 1 and 2, two multiple-regression analyses were computed in order to evaluate the potential role of confounded variables. Item latencies were entered in the regressions as a dependent variable and word frequency, summed bigram frequency, orthographic neighborhood density, number of higher frequency neighbors, phonological neighborhood density, and item delayed latencies as independent variables. When number of phonemes was added to the independent variables, all these variables explained 39.9% of the variance [$F(7,34) = 3.23, p < .01$], but only number of phonemes and item delayed latencies accounted for a unique and significant part of the variance [$t(34) = 2.2$, $p < .05$ and $t(34) = 3.29, p < .01$, respectively]. Similarly, when regularity was entered in the list of independent variables, these variables explained 41.9% of the variance [$F(7,34) = 3.5, p < .01$], but, again, only regularity and item delayed latencies accounted for a unique and significant part of the variance [$t(34) = 2.47, p < .05$ and $t(34) = 3.26, p < .01$, respectively].

Discussion

The main finding of Experiment 3 is that we replicate the same pattern of results as in Experiment 2, using a different experimental paradigm; that is, a naming task. Words with multiletter graphemes are responded to faster than words consisting of single-letter graphemes only. Regular words are responded to faster than irregular ones. These effects of graphemic complexity and multiple print-to-sound associations are observed jointly and appear to be cumulative.

These results are consistent with our prediction, since we assumed that these effects occur before the level of phonological output and therefore should affect perceptual identification and naming latencies in similar ways. However, it has been shown that naming latencies are also strongly determined by variables such as the type of initial phoneme. For example, Spieler and Balota (1997) found that this phonological output factor explained 29.9% of the variance in a large-scale naming experiment. Conducting a delayed naming task immediately after the naming task allowed us to control for this factor since our stimuli were not initially matched on that dimension. When delayed naming latencies were entered as a covariate in the ANOVAs, all our effects reached significance both for participants and items. This result indicates that, without a model taking into account phonological output factors and determining their influence in the time course of naming, delayed naming latencies provide a good means to control for this source of variance that has not been accounted for so far by current models of word reading.

GENERAL DISCUSSION

The main finding of Experiments 1–3 concerns the independence of the graphemic complexity and the multiple print-to-sound associations effects. Experiment 1 shows that the graphemic complexity effect is observed when graphemic regularity and body consistency are controlled. This effect, therefore, has an autonomous status and is obtained independently of multiple-associations effects. Experiment 2 provides further empirical evidence in favor of the independence of the two effects since both are jointly observed in the same experiment. Experiment 3 replicates and extends these results to the naming task, indicating that graphemic complexity and multiple print-to-sound associations affect processing levels that are shared by perceptual identification and naming.

An original characteristic of the present data concerns the experimental paradigm in which they have been observed. Contrary to the classical naming task, the luminance-increasing paradigm does not require participants to produce an overt phonological response. Participants always give the same simple motor response by pressing the space bar as soon as they have identified the target word but do not have to pronounce it. This does not mean, however, that no phonological information is required for performing this task. Indeed, just after pressing the space bar, the target word disappears from the screen, and participants have to enter the word they have identified, using the keyboard. This procedure could be done only on the basis of the visual and orthographic short-term memory trace of the word, but more probably, participants recode it phonologically. Phonology is therefore very likely involved at that point.

This has a major consequence for modeling reading processes and human performance in the present paradigm. As in the naming task, both orthographic and phonological processes are required and coactivated in the luminance-increasing paradigm in order to identify the target word. But, contrary to the naming task, the fact that no articulation response is required rules out any explanation of the observed effects in terms of phonological output or motor programming differences. This is indeed a serious limitation of the naming task since many studies have demonstrated recently that factors related to the pronunciation of words explain a large amount of variance in naming (e.g., Kessler et al., 2002; Spieler & Balota, 1997; Treiman et al., 1995). One advantage of the present results is to show that graphemic complexity and regularity effects take place at earlier processing stages than at stages related to the articulatory-motor programming of the word.

Another critical feature of these experiments is the fact that the manipulation of graphemic complexity affected the processing of words in our experiments (see also Rey et al., 1998) and not only nonwords (e.g., Rastle & Coltheart, 1998). This obviously imposes new constraints for modeling. For example, within the framework of the dual-route model (Coltheart et al., 2001; Rastle & Coltheart, 1998), the graphemic complexity effect was originally located within the nonlexical route (the processing route that converts any letter sequence into a sequence of phonemes by applying a set of grapheme-to-phoneme conversion rules, i.e., the most frequent grapheme-to-phoneme associations; Coltheart et al., 1993; Coltheart et al., 2001). According to this model, the graphemic complexity effect would be a consequence of the serial letter-by-letter processing performed within the nonlexical route (for a detailed description of this account, see Rastle & Coltheart, 1998). This explanation concerned, however, only the processing of nonwords, and it remains to be seen whether or not it could be extended to words since, in this model, words are processed within the lexical route where the orthographic representation of words is stored and accessed directly from the letter level.

One possible solution is provided by another assumption of the model, stipulating that the two routes are working in parallel and that processing of low-frequency words can be influenced by processing conflicts within the nonlexical route (this is, in fact, the way the model accounts for regularity effect). This assumption could explain why the words used in our experiments display a graphemic complexity effect since we used words of low/medium frequency (words in Experiment 1 had a mean frequency of 7 occurrences per million, while words in Experiments 2 and 3 had a mean frequency of 12 occurrences per million). Given that low-frequency words are supposed to be accessed less rapidly in the lexical route, their processing might therefore be slowed down by the conflicts generated in the nonlexical route with multiletter graphemes. Computer simulations need to be done, however, in order to see whether the processing dynamics of the dual-route model can indeed handle this graphemic effect on words together with the regularity effect (in order to account for the results of Experiments 2 and 3).

The present empirical data therefore theoretically support the processing assumptions of the dual-route model. But are they also consistent with alternative computational architectures of reading like the ones proposed by learning models using the backpropagation algorithm (e.g., Plaut et al., 1996; Zorzi et al., 1998)? Although computer simulations would also be necessary to confirm our claims, we believe that both of these computational models can generate the present effects due to some of their structural properties. In both models, during the training phase, the networks indeed learn to establish associations from orthographic units (which are slightly different in each model, but these orthographic coding choices are not crucial for our demonstration) to phonemic units, and this knowledge is encoded within connection weights. For example, both models learn to associate a grapheme like EA with its corresponding phonemic pronunciations in words like BREAD, BREAK, BEACH, and so forth. EA is more frequently associated with one of these pronunciations (i.e., the "regular" one). When a less frequent pronunciation is expected, the networks will manage to produce the correct phoneme (when learning of all lexical entries has been completed) but the more frequently associated phoneme will also be highly activated (or the probability to produce it will be very high). If one assumes a response mechanism based on the relative activation of each phoneme, a difference between regular and irregular words will obviously emerge. Although we acknowledge that this verbal description should be complemented by a detailed computational account in terms of processing dynamics, rate of learning, or response accuracy, it indicates that these learning networks have inherent problems in establishing and activating irregular grapheme-to-phoneme associations. Similarly, as they learn to associate the multiletter grapheme EA with its corresponding possible phonemes, the networks also encode associations between single-letter graphemes E and A and other phonemes in words like FRESH or GRASS, for example. Here again, after the training phase, when the letters E and A are presented together, they will activate not only phonemes associated to the multiletter grapheme EA but also phonemes associated to the single-letter graphemes E and A. This multiple phonemic activation can probably account for the graphemic complexity effect because words with multiletter graphemes generate more competing phonemic activation than words essentially composed of single-letter graphemes.

Although both the regularity and graphemic complexity effects seem to be predicted by dual-route and learning models of reading, the present experiments provide further arguments concerning the debate on the so-called consistency effect for which the two categories of models have different predictions. As indicated in the introduction, the problem of multiple associations between print and sound has been addressed in two different ways in the experimental literature. On the one side, Glushko (1979) and more recently Jared (1997, 2002) have observed that naming latencies were influenced by print-to-sound consistency of orthographic bodies. It follows from this view that orthographic bodies must have a special status within the reading system or at least that a model of reading should integrate such an orthographic unit at some level of processing. This is explicitly the case in the model of Zorzi et al. (1998), which assumes a segmentation mechanism that separates a word's onset from its orthographic body before spreading activation from orthographic nodes to phonological nodes. In the Plaut et al. (1996) model, a similar segmentation mechanism is also assumed since orthographic units are organized following an onset–nucleus–coda scheme at the orthographic level. Both models can therefore predict consistency effects because of these segmentation assumptions. For the dual-route model, however, graphemes are the unique orthographic units that participate in print-to-sound processes, and orthographic bodies do not have any status within the model. Demonstrating that orthographic bodies' consistency influences naming latencies would therefore falsify some of the assumptions of the dual-route model. However, what empirical evidence so far supports such a view?

The major challenge was to observe a consistency effect while regularity is controlled. According to Coltheart et al. (2001), there exists one clear description of such an effect by Jared (1997; see also Cortese & Simpson, 2000; Jared, 2002). However, when these data were compared with computer simulations of the dual-route model, it turned out that the model was in fact able to capture the consistency effect. A detailed investigation of the model's dynamic revealed that the alleged consistency effect was generated within the nonlexical route and could be attributed to the presence of multiletter graphemes (called whammies in Coltheart et al., 2001, and Rastle & Coltheart, 1998). Orthographic body consistency was indeed confounded with graphemic complexity. Of course, although these simulations definitely provide an alternative explanation of the results, the reported effects could nevertheless still be due to body consistency and not graphemic complexity. Experiments manipulating consistency, while controlling for graphemic complexity on the one hand and manipulating graphemic complexity while controlling for consistency on the other, will help to disentangle these alternative explanations.

In the present study, one part of the solution has been provided since all words were selected as consistent (i.e., their orthographic body being always pronounced in the same way). In this case, a graphemic complexity effect was repeatedly observed, and it could not be attributed to body consistency. These data therefore show that the graphemic complexity effect is independent from the consistency effect. However, in order to give another chance to the orthographic body hypothesis, we looked at the number and frequency of friends—that is, the number and frequency of words sharing the same rime—a variable that is closer to the one manipulated by Jared (1997, 2002; although, to be precise and according to Jared, it is the relative strength of friends and enemies that has been shown to affect naming latencies, response times being longer for words with few friends and many enemies).

The hypothesis of a confounded factor related to orthographic bodies—namely, the number or the frequency of friends—was tested with data from Experiment 1. We used the database computed by Ziegler, Stone, and Jacobs (1997) providing for each English rime the number and summed frequency of words sharing that rime (i.e., friends), to calculate these values for each target word. These analyses revealed that, on average, the number and frequency of friends was greater for five-phoneme words compared with three-phoneme words. The difference between the two experimental groups was significant for the number of friends [$t(28) = 4.3, p < .001$] and marginally significant for the summed frequency of friends [$t(28) = 1.83, p = .07$]. These factors appeared therefore to be confounded with our experimental manipulation.

However, this hypothesis did not receive support from a multiple-regression analysis that we conducted on our dataset. Word identification latencies from Experiment 1 were entered as a dependent variable, and the number of phonemes (i.e., three and five), the number of friends, and the summed frequency of friends were entered as independent variables. These three factors together significantly explained 28% of the variance [$F(3,26) = 3.34$, $p < .05$], but among these factors only one, the number of phonemes, accounted for a unique and significant part of the variance [$t(26) = 2.15, p < .05$]. Neither the number of friends nor the summed frequency of friends reached significance (all ts < 1). The conclusion of these analyses is that, although body-related factors such as number of friends and summed frequency of friends were found to be confounded with our experimental manipulation, these factors did not account for a unique and significant part of the variance, nor did they cancel the effect of our experimental manipulations. We can therefore confidently rule out an alternative interpretation of our results in terms of body-related factors.

To conclude, the present study describes an empirical clarification concerning the graphemic complexity effect reported recently (Rastle & Coltheart, 1998; Rey et al., 1998). The results of our experiments allow us to rule out any interpretation of this effect in terms of multiple print-to-sound associations effects (i.e., regularity or consistency effects). The effects of graphemic complexity and regularity were even shown to be cumulative, again indicating their independence. Finally, while the results show that a graphemic complexity effect can be observed while controlling for orthographic body-related factors (consistency, number, or summed frequency of friends), it remains to be seen whether or not previously reported effects of body consistency can still be observed while controlling for graphemic complexity and regularity.

REFERENCES

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM]. Linguistic Data Consortium, University of Pennsylvania, Philadelphia.

Cohen, J. [D.], MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, **25**, 257-271.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151-216). London: Academic Press.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, **100**, 589-608.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual-route cascaded model of visual word recognition and reading aloud. *Psychological Review*, **108**, 204-256.

Cortese, M. J., & Simpson, G. B. (2000). Regularity effects in word naming: What are they? *Memory & Cognition*, **28**, 1269-1276.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 674-691.

Jacobs, A. M., Rey, A., Ziegler, J. C., & Grainger, J. (1998). MROM-P: An interactive activation, multiple read-out model of orthographic and phonological processes in visual word recognition. In J. Grainger & A. M. Jacobs (Eds.), *Symbolic connectionist approaches to human cognition* (pp. 147-188). Hillsdale, NJ: Erlbaum.

Jared, D. (1997). Spelling–sound consistency affects the naming of high-frequency words. *Journal of Memory & Language*, **36**, 505-529.

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory & Language*, **46**, 723-750.

Kenstowicz, M. (1994). *Phonology in generative grammar*. Cambridge, MA: Blackwell.

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory & Language*, **47**, 145-171.

Lange, M. (2002). Activation of multiple phonemic associates of graphemes in visual word recognition. *Brain & Language*, **80**, 610-620.

Norris, D. (1994). A quantitative multiple-levels model of reading aloud. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 1212-1232.

Peereman, R., Brand, M., & Rey, A. (in press). Letter-by-letter processing in the phonological conversion of multi-letter graphemes: Searching sounds in printed pseudowords. *Psychonomic Bulletin & Review*.

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.

Rastle, K., & Coltheart, M. (1998). Whammies and double whammies: The length effect on nonword reading. *Psychonomic Bulletin & Review*, **5**, 277-282.

Rey, A., Jacobs, A. M., Schmidt-Weigand, F., & Ziegler, J. C. (1998). A phoneme effect in visual word recognition. *Cognition*, **68**, B71-B80.

Rey, A., Ziegler, J. C., & Jacobs, A. M. (2000). Graphemes are perceptual reading units. *Cognition*, **75**, B1-B12.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning & Verbal Behavior*, **23**, 383-404.

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

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory & Language*, **26**, 608-631.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, **124**, 107-136.

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What's the pronunciation for *-ough* and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, & Computers*, **29**, 600-618.

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 & Performance*, **24**, 1131-1161.

APPENDIX A
Five-Letter Words Used in Experiment 1

All words are regular (following Coltheart's [1978] definition) and feedforward consistent.

Three-Phoneme Words
beech, booze, feign, heave, leech, loath, maize, peach, poach, pouch, reign, roach, teeth, waive, weave

Five-Phoneme Words
blink, blunt, brisk, clank, cleft, crest, crisp, crust, frank, frond, grunt, spank, stomp, trump, tract

APPENDIX B

All the words used in Experiments 2 and 3 were feedforward consistent. The three-phoneme irregular words were *bathe*, *guise*, *lathe*, *mauve*, *niche*, *pearl*, *route*, *seize*, *sieve*, *thief*, *thyme*, *vague*, *weird*, *yearn*; the three-phoneme regular words were *baulk*, *beard*, *birch*, *chain*, *churn*, *hoard*, *kneel*, *lurch*, *mourn*, *niece*, *quirk*, *shawl*, *shoal*, *whirl*; the five-phoneme regular words were *blunt*, *brisk*, *brunt*, *cleft*, *crisp*, *drift*, *drops*, *frond*, *skulk*, *spank*, *steps*, *stunt*, *trump, twist.*

Table B1
Mean Values for Independent Variables in Experiments 2 and 3

Words	Word Frequency	Summed Bigram Frequency	Orthographic Neighborhood Density	No. Higher Frequency Neighbors	Phonological Neighborhood Density
Three-phoneme irregular	13	4,846	0.79	0.50	0.50
Three-phoneme regular	12.3	4,812	0.79	0.64	0.64
Five-phoneme regular	13	4,036	1.07	0.64	0.57

(Manuscript received April 17, 2003;
revision accepted for publication May 6, 2004.)