

BRIEF REPORTS

The influence of spelling on phonological encoding in word reading, object naming, and word generation

ARDI ROELOFS

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Does the spelling of a word mandatorily constrain spoken word production, or does it do so only when spelling is relevant for the production task at hand? Damian and Bowers (2003) reported spelling effects in spoken word production in English using a prompt–response word generation task. Preparation of the response words was disrupted when the responses shared initial phonemes that differed in spelling, suggesting that spelling constrains speech production mandatorily. The present experiments, conducted in Dutch, tested for spelling effects using word production tasks in which spelling was clearly relevant (oral reading in Experiment 1) or irrelevant (object naming and word generation in Experiments 2 and 3, respectively). Response preparation was disrupted by spelling inconsistency only with the word reading, suggesting that the spelling of a word constrains spoken word production in Dutch only when it is relevant for the word production task at hand.

An important issue in psychology in general, and psycholinguistics in particular, is the extent to which cognitive processes operate in a mandatory fashion. For example, does the spelling of a word necessarily constrain spoken word production, or does it do so only when spelling is relevant for the production task at hand? Whereas orthographic processing is relevant for word reading, it is not directly relevant for object naming. If the spelling of a word constrains spoken word production regardless of its relevance, effects of spelling should be obtained both in reading and object naming. In contrast, if the spelling constrains word production only when it is relevant for the task at hand, effects of spelling should be obtained in reading but not in object naming. Recently, Damian and Bowers (2003) reported evidence suggesting mandatory influences of spelling on spoken word production. This evidence was obtained using a prompt–response word generation task and the preparation paradigm developed by Meyer (1990, 1991). In this section, I will first explain the preparation paradigm and the relevant findings, before discussing the findings of Damian and Bowers.

This research was supported by the Max Planck Institute for Psycholinguistics, the F. C. Donders Centre for Cognitive Neuroimaging, and the Nijmegen Institute for Cognition and Information. I am indebted to Sascha Oberrecht and Philipp Rauch for their help in preparing and running the experiments, and to Markus Damian, Derek Besner, the members of the Utterance Encoding Project of the Max Planck Institute for Psycholinguistics, and an anonymous reviewer for helpful comments. Correspondence relating to this article may be sent to A. Roelofs, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands (e-mail: ardi@mpi.nl).

Meyer's (1990, 1991) preparation paradigm falls into the class of action preparation and precuing techniques that have been widely used in studying the planning of skilled action. For example, Rosenbaum (1980) used precuing to control the amount of preparation in arm movement. He manipulated the uncertainty in the specification of arm direction and extent, and observed that as more information was available in advance to allow preparation, the movement onset latency decreased. To study the effect of advance phonological information on word preparation, Meyer (1990, 1991) asked participants to generate spoken words in response to different, written prompt words. Dutch participants first learned small sets of prompt–response pairs such as FRUIT–“melon,” IRON–“metal,” GRASS–“meadow” (the examples are in English, but the experiments were conducted in Dutch). During the following test phase, they had to orally produce the response word of a pair (e.g., “melon”) upon visual presentation on a computer screen of the prompt word (FRUIT). On each trial, one of the prompts was presented. The order of prompts across trials was random. The word production latency—the interval between prompt onset and speech onset—was the main dependent variable. Each experiment included homogeneous and heterogeneous sets. In homogeneous sets, the response words shared a part of their form, such as the first syllable in “melon,” “metal,” “meadow,” or the second syllable in “pocket,” “ticket,” “bucket.” In heterogeneous sets, the response words were unrelated in form. Regrouping the pairs from the homogeneous sets created the heterogeneous sets.

Meyer found shorter word production latencies in homogeneous than in heterogeneous sets, henceforth re-

ferred to as the *preparation effect*. This preparation effect was obtained only when the response words in homogeneous sets shared word-initial phonemes (as in “melon,” “metal,” “meadow”), not when they shared word-final phonemes (as in “pocket,” “ticket,” “bucket”). The magnitude of the preparation effect increased with the number of shared phonemes. Subsequent research in Dutch showed that the magnitude of the preparation effect also depends on morpheme structure and other abstract linguistic variables (see, e.g., Janssen, Roelofs, & Levelt, 2002; Roelofs, 1996b; Roelofs & Meyer, 1998). For example, the preparation effect was larger when the shared phonemes made up a morpheme in the response words than when the same phonemes did not make up a morpheme (Roelofs, 1996b; Roelofs & Baayen, 2002), and it was also larger for low-frequency than for high-frequency morphemes (Roelofs, 1996a, 1998). Roelofs (1999, 2003) observed that preparing phonological features alone does not facilitate spoken word production. When disyllabic words were produced in sets that differed in form or that either shared the first syllable (e.g., “baby,” “bagel,” “baker”) or shared the first syllable except for a single phonological feature in the initial phoneme (e.g., “data,” “baby,” and “bagel,” in which the syllables /dej/ and /bej/ share all phonological features, except that the initial consonants /d/ and /b/ differ in place of articulation), only fully shared first syllables yielded a preparation effect. The special status of phonemic identity suggests that preparation happens at the level of phonemes rather than of phonological features. Roelofs (1999, 2003, 2004) showed that the preparation effect is not only obtained with prompt–response word generation, but also with naming objects and reading their names. A computational account of the preparation effect in terms of partial planning of speech was given by Roelofs (1997a, 1997b; Levelt, Roelofs, & Meyer, 1999) using WEAVER++.

Earlier research provided some evidence that the preparation effect occurs regardless of whether the spelling is shared or not. Meyer (1990) observed preparation effects for shared syllables (i.e., /sij/) of 42 msec (her Experiment 1) and 49 msec (her Experiment 3) with varying orthography (i.e., *ci* versus *si*). Moreover, Chen, Chen, and Dell (2002) demonstrated in Mandarin Chinese that preparation of syllables is not dependent on shared orthography, by using two-syllable two-character response words. When the first syllable was shared but not the first character, the preparation effect was as large as when syllable and character were both shared.

However, in experiments conducted in English, Damian and Bowers (2003) observed that when the words in a set shared initial phonemes that differed in spelling, no preparation effect was obtained (their Experiments 1 and 2). For example, a preparation effect was obtained for “camel,” “coffee,” “cushion,” but not for “kennel,” “coffee,” “cushion.” The preparation effect was absent even when the prompt–response pairs were learned auditorily and the prompts were spoken words (their Experiment 3). Finally, when the first letter was shared but not the initial phoneme (their Experiment 4), as in “census,” “climate,”

“candle,” the preparation effect was also absent. Thus, phonological planning is a necessary but not sufficient condition for obtaining the preparation effect.

Damian and Bowers (2003) took their findings as evidence for the view that the spelling of a word constrains spoken word production regardless of its relevance. They conjectured that “sounds and letters have to enter a stable state of congruency to be selected or planned. Hence, in inconsistent blocks, due to the shared initial segments of the stimuli, the correct phoneme can be planned, but the correct letter remains ambiguous. If it takes some time to resolve the orthographic ambiguity, then the priming effect due to the phonological planning could be reduced or eliminated” (p. 129). It is unclear, however, whether planning in terms of both phonemes and letters is mandatory or whether it is a strategy. Perhaps Damian and Bowers’s participants encoded the initial sounds of the responses in both orthographic and phonological forms to help maintain them in memory, thereby hampering response preparation in inconsistent sets.

The present experiments examined the spelling disruption effect in Dutch with tasks in which the spelling of a word was relevant (namely, oral reading; Experiment 1) or irrelevant (object naming; Experiment 2), as well as with prompt–response word generation (Experiment 3), the task used by Damian and Bowers (2003) and Meyer (1990, 1991). If the spelling disruption effect is obtained in reading but not in object naming or word generation, this would suggest that the disruption happens, at least for Dutch, only when spelling is relevant and is used in the task at hand.

EXPERIMENT 1

Reading

Method

Participants. The experiment was conducted with 18 paid participants from the pool of the Max Planck Institute. They were native speakers of Dutch.

Materials and Design. The stimuli consisted of 12 written words (Experiment 1), corresponding pictures (Experiment 2), and written words that had the response words as associates (Experiment 3). The response words were disyllabic nouns with stress on the second syllable. The initial phoneme of six response words was /s/, and for the other six responses it was /k/. The initial /s/ was spelled *s* in half of the cases and *c* in the other half. The initial /k/ was spelled *k* in half of the cases and *c* in the other half. The pictures were line drawings of simple objects from the picture gallery available at the Max Planck Institute. The pictures were digitized and scaled to fit into a virtual frame of 10 × 10 cm. The words were presented in 36-point lowercase Arial font and were grouped into 12 response sets of three stimuli each (see Table 1). Each set was presented in a separate block of trials. In four *homogeneous* sets, the response words shared the initial phoneme and letter. In four *inconsistent* sets, all three response words shared the initial phoneme, but only two responses shared the initial letter. In the remaining four *heterogeneous* sets, the response words had varying initial phonemes and letters. This variable—homogeneous versus inconsistent versus heterogeneous sets—will be called *context*. Each written word was presented in all three conditions; only their combinations into sets differed.

Each participant was tested once on each set, and each of the words in a set was tested 10 times within a block of trials. The order

Table 1
Response Sets of Experiments 1–3

Context	Set
Homogeneous	1. citroen, circuit, CD (lemon, circuit, CD)
	2. sandaal, sigaar, soldaat (sandal, cigar, soldier)
	3. contract, colbert, cadeau (contract, jacket, present)
	4. kompas, kanon, konijn (compass, cannon, rabbit)
Inconsistent	5. sandaal, circuit, CD
	6. citroen, sigaar, soldaat
	7. kompas, colbert, cadeau
	8. contract, kanon, konijn
Heterogeneous	9. citroen, circuit, konijn
	10. sandaal, sigaar, cadeau
	11. contract, colbert, soldaat
	12. kompas, kanon, CD

of testing the words was random, except that immediate repetitions of stimuli were excluded. The sets were presented in “superblocks” so that four consecutive sets always presented the same type of context (i.e., homogeneous, inconsistent, or heterogeneous). The order in which the participants received the superblocks was such that each superblock was tested in each serial position an equal number of times. The order in which each set was tested in a superblock was sequentially balanced.

Procedure and Apparatus. The participants were tested individually. They were seated in a quiet room in front of a computer screen (NEC Multisync 30) and a microphone (Sennheiser ME40) connected to an electronic voice key. The distance from the screen was approximately 50 cm. Before the experiment, a participant received written instructions about the task. Before each trial block, the three words of a set were presented on the screen. Thus, there was foreknowledge about the form of the reading responses before the first trial of a block.

The structure of a trial was as follows. A trial was started by the display of a word, in white on a black background, for 1.0 sec. Before the start of the next trial, there was a blank interval of 1.5 sec. Thus, the total duration of a trial was 2.5 sec. A Hermac computer controlled stimulus presentation and data collection, including the voice key.

Analyses. The response coding and analyses were the same in all experiments. After each trial, the experimenter coded the responses for errors. Five types of incorrect responses were distinguished: wrong response words, wrong pronunciation of the words, disfluencies, triggering of the voice key by nonspeech sounds, and failures to respond within 1.5 sec after word presentation. Incorrect responses were excluded from the statistical analysis of the production latencies. For all experiments, by-participants (F_1) and by-items (F_2) ANOVAs were performed on the latencies and error rates.

Results and Discussion

Table 2 gives the mean word reading latencies, their standard deviations, and the error percentages for Experiment 1. A sizable preparation effect of 38 msec was obtained when the spelling was shared in a set (the homoge-

neous condition), but only a small effect of 11 msec when the spelling varied (the inconsistent condition).

The analysis of the reading latencies yielded a main effect of context [$F_1(2,34) = 5.83, MS_e = 2,437, p < .007; F_2(2,20) = 33.09, MS_e = 126, p < .001$]. Planned comparisons showed that the latencies were shorter in the homogeneous than in the heterogeneous sets [$t_1(17) = 3.2, p < .003; t_2(11) = 8.1, p < .001$] and shorter in the homogeneous than in the inconsistent sets [$t_1(17) = 2.1, p < .03; t_2(11) = 7.2, p < .001$], but the latencies did not differ reliably between the inconsistent and heterogeneous sets [$t_1(17) = 1.2, p = .13; t_2(11) = 2.4, p < .04$]. The analysis of the reading errors yielded no significant results.

To summarize, in oral reading, participants could benefit from foreknowledge of the initial phoneme of a word, but only when the spelling of the phoneme was also shared. The data for reading replicate the findings that Damian and Bowers (2003) obtained for prompt–response word generation.

EXPERIMENT 2
Object Naming

Method

Participants. The experiment was conducted with 12 paid Dutch participants from the pool of the Max Planck Institute. None of them took part in Experiment 1.

Materials, Design, Procedure, Apparatus, and Analyses. These were the same as in Experiment 1, except that pictured objects were presented rather than written words. Before each block of trials, the participants were shown the three pictures of a set, thereby providing foreknowledge about the form of the responses before the first trial of a block.

Table 2
Mean Response Latencies (*M*, in Milliseconds), Standard Deviations (*SD*), and Error Percentages (*E%*) per Context and Task in Experiments 1–3

Task	Context								
	Homogeneous			Inconsistent			Heterogeneous		
	<i>M</i>	<i>SD</i>	<i>E%</i>	<i>M</i>	<i>SD</i>	<i>E%</i>	<i>M</i>	<i>SD</i>	<i>E%</i>
Reading	450	96	1.2	477	103	1.2	488	104	1.4
Object naming	536	139	1.8	535	142	1.3	567	142	1.7
Word generation	661	165	2.7	656	164	2.2	682	175	2.3

Results and Discussion

Table 2 also gives the mean object naming latencies, their standard deviations, and the error percentages for Experiment 2. The table shows that a preparation effect was obtained regardless of whether the spelling was shared (31 msec for the homogeneous condition) or not shared (32 msec for the inconsistent condition).

The analysis of the naming latencies yielded a marginally significant main effect of context [$F_1(2,22) = 3.14$, $MS_e = 2,530$, $p < .06$; $F_2(2,20) = 19.61$, $MS_e = 174$, $p < .001$]. Planned comparisons showed that the latencies were shorter in the homogeneous than in the heterogeneous sets [$t_1(11) = 1.8$, $p < .05$; $t_2(11) = 6.7$, $p < .001$] and shorter in the inconsistent than in the homogeneous sets [$t_1(11) = 3.31$, $p < .004$; $t_2(11) = 8.5$, $p < .001$], but the latencies did not differ between the homogeneous and inconsistent sets [$t_1(11) = 0.01$; $t_2(11) = 0.1$, $p = .93$]. The analysis of the object naming errors yielded no significant results.

To summarize, in object naming, participants could benefit from foreknowledge of the initial phoneme of a word, regardless of whether the spelling of the phoneme was also shared. Thus, the spelling disruption that was obtained with word reading (Experiment 1) did not occur with object naming. This suggests that the disruption arises only when the spelling of a word is task relevant.

EXPERIMENT 3 Word Generation

Method

Participants. The experiment was conducted with 12 paid Dutch participants from the pool of the Max Planck Institute. None of them took part in Experiments 1 and 2.

Materials, Design, Procedure, Apparatus, and Analyses. These were the same as in Experiments 1 and 2. Participants were again tested on all 12 sets (Table 1). However, there were now no reading responses (Experiment 1) or picture naming responses (Experiment 2). Instead, the participants first learned the words of the sets as response members of prompt–response pairs and then produced a response when presented with the corresponding prompt, exactly as in the experiments of Damian and Bowers (2003) and Meyer (1990, 1991). Before each block of trials, the three pairs of a set were shown on the screen. As soon as the participant indicated having studied the pairs sufficiently, the experimenter started the block of test trials. On each trial, one of the prompts was presented. The order of prompts across trials was random.

Results and Discussion

Table 2 also gives the mean word generation latencies, their standard deviations, and the error percentages for Experiment 3. The table shows that a preparation effect was obtained regardless of whether the spelling was shared (21 msec for the homogeneous condition) or not (26 msec for the inconsistent condition).

The analysis of the word generation latencies yielded a main effect of context [$F_1(2,22) = 3.81$, $MS_e = 1,393$, $p < .04$; $F_2(2,20) = 11.76$, $MS_e = 218$, $p < .001$]. Planned comparisons showed that the latencies were shorter in the homogeneous than in the heterogeneous sets [$t_1(11) = 2.7$, $p < .01$; $t_2(11) = 3.1$, $p < .01$] and shorter in the inconsis-

tent than in the heterogeneous sets [$t_1(11) = 2.2$, $p < .03$; $t_2(11) = 3.8$, $p < .003$], but the latencies did not differ between the homogeneous and inconsistent sets [$t_1(11) = 0.3$, $p = .76$; $t_2(11) = 1.1$, $p = .28$]. The analysis of the word generation errors yielded no significant results.

To summarize, in word generation, participants could benefit from foreknowledge of the initial phoneme of a word, regardless of whether the spelling of the phoneme was also shared. Thus, the disruption that Damian and Bowers (2003) obtained with prompt–response word generation was not replicated. In the present experiment, there was a nonsignificant 5-msec difference in preparation effect between the homogeneous and the inconsistent sets. However, if anything, the preparation effect was numerically larger for the inconsistent sets than for the homogeneous sets, unlike the results Damian and Bowers obtained.

GENERAL DISCUSSION

The aim of the research reported in this article was to examine the influence of spelling on phonological encoding in spoken word production. Damian and Bowers (2003) reported effects of spelling in speech production using a prompt–response word generation task. Spoken word preparation was disrupted when the response words in a set of paired associates shared initial phonemes that differed in spelling. The present experiments tested for spelling effects using word production tasks in which spelling was relevant (oral reading in Experiment 1) or irrelevant (picture naming and word generation in Experiments 2 and 3, respectively). Response preparation was disrupted by spelling inconsistency only with reading (Experiment 1), suggesting that the spelling of a word constrains word production only when it is relevant for the task at hand.

Damian and Bowers (2003) argued that their observation of spelling disruption demanded that earlier findings obtained with the preparation paradigm should be reinterpreted. In particular, Roelofs (1999, 2003) observed that preparing phonological features alone does not facilitate spoken word production. When disyllabic words were produced in sets different in form, or in sets sharing either the first phoneme or the first phoneme except for a single phonological feature, only fully shared first phonemes yielded facilitation. The special status of phonemic identity suggests that preparation happens in terms of phonemes, not of phonological features. However, given the results of Damian and Bowers, the special status of phonemic identity might have been a spelling effect. When initial sounds share phonological features only, their spelling will often differ. If spelling always constrains word production, this would explain the absence of a preparation effect in Roelofs (1999, 2003). However, Roelofs (1999) observed that the effect of phonemic identity is also obtained in object naming. The present Experiment 2 suggests that a difference in spelling does not disrupt the preparation effect in object naming. Moreover, the present Experiment 3 suggests that spelling does not necessarily

disrupt preparation in the prompt–response word generation task, either. Thus, there is no need to reinterpret the findings of Roelofs (1999, 2003).

The question remains why Damian and Bowers (2003) obtained an effect of spelling in the prompt–response word generation task (they did not test object naming and reading), whereas in the present Experiment 3, using the same task, no such effect was observed. It is unlikely that methodological differences between Damian and Bowers and the present Experiment 3 caused the difference in outcomes, because the methods were almost identical (same design, same number of responses in a set, similar number of item repetitions within a trial block, etc.). Moreover, the response latencies were very similar for Damian and Bowers's experiments (overall mean: 668 msec) and for the present Experiment 3 (overall mean: 666 msec). However, a salient difference between the studies was language: Participants in the Damian and Bowers experiments were English speakers, whereas those in the present experiments were Dutch. It is possible that cross-linguistic differences exist in the degree to which orthography and phonology interact in speech production, perhaps related to differences in orthographic depth between languages (Damian & Bowers, 2003).

Another possibility is a difference in strategic use of orthographic information. Visual images provide a very powerful way of associating items (Baddeley, 1997). All responses in the present Experiment 3 referred to concrete and imageable objects, whereas this was not the case in the experiments of Damian and Bowers (2003). Perhaps Damian and Bowers's participants encoded the initial sounds of the responses in both orthographic and phonological forms to help maintain them in memory, whereas the participants in the present Experiment 3 encoded them only phonologically.

A strategic explanation has also been proposed for the influence of spelling on speech perception tasks. For example, Dijkstra, Roelofs, and Fieuws (1995) observed that in a phoneme monitoring task performed on Dutch spoken words, the response latencies were affected by the spelling of the words. Cutler, Treiman, and van Ooijen (1998), on the other hand, showed that the effect of spelling on phoneme monitoring disappeared when a large number of consistently spelled filler items diverted participants' attention from the spelling of the words, suggesting a strategic origin.

To conclude, the reported experiments addressed the issue of whether the spelling of a word constrains spoken word production only in reading or also in object naming and word generation. An effect of spelling was obtained only in reading, suggesting that the spelling of a word in Dutch plays a role only when it is relevant for the task at hand.

REFERENCES

- BADDELEY, A. D. (1997). *Human memory: Theory and practice*. Hove, U.K.: Psychology Press.
- CHEN, J.-Y., CHEN, T.-M., & DELL, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory & Language*, **46**, 751-781.
- CUTLER, A., TREIMAN, R., & VAN OOIJEN, B. (1998). Orthografik inkonstistency ephekts in foneme detektion? In R. H. Mannell & J. Robert-Ribes (Eds.), *Proceedings of the Fifth International Conference on Spoken Language Processing* (pp. 2783-2786). Sydney: ICSLP.
- DAMIAN, M. F., & BOWERS, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory & Language*, **49**, 119-132.
- DIJKSTRA, T., ROELOFS, A., & FIEUWS, S. (1995). Orthographic effects on phoneme monitoring. *Canadian Journal of Experimental Psychology*, **49**, 264-271.
- JANSSEN, D. P., ROELOFS, A., & LEVELT, W. J. M. (2002). Inflectional frames in language production. *Language & Cognitive Processes*, **17**, 209-236.
- LEVELT, W. J. M., ROELOFS, A., & MEYER, A. S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences*, **22**, 1-38.
- MEYER, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory & Language*, **29**, 524-545.
- MEYER, A. S. (1991). The time course of phonological encoding in language production: The phonological encoding inside a syllable. *Journal of Memory & Language*, **30**, 69-89.
- ROELOFS, A. (1996a). Morpheme frequency in speech production: Testing WEAVER. In G. E. Booij & J. van Marle (Eds.), *Yearbook of morphology 1996* (pp. 135-154). Dordrecht: Kluwer.
- ROELOFS, A. (1996b). Serial order in planning the production of successive morphemes of a word. *Journal of Memory & Language*, **35**, 854-876.
- ROELOFS, A. (1997a). Syllabification in speech production: Evaluation of WEAVER. *Language & Cognitive Processes*, **12**, 659-696.
- ROELOFS, A. (1997b). The WEAVER model of word-form encoding in speech production. *Cognition*, **64**, 249-284.
- ROELOFS, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: Verb–particle combinations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 904-921.
- ROELOFS, A. (1999). Phonological segments and features as planning units in speech production. *Language & Cognitive Processes*, **14**, 173-200.
- ROELOFS, A. (2003). Shared phonological encoding processes and representations of languages in bilingual speakers. *Language & Cognitive Processes*, **18**, 175-204.
- ROELOFS, A. (2004). Seriality of phonological encoding in naming objects and reading their names. *Memory & Cognition*, **32**, 212-222.
- ROELOFS, A., & BAAYEN, H. (2002). Morphology by itself in planning the production of spoken words. *Psychonomic Bulletin & Review*, **9**, 132-138.
- ROELOFS, A., & MEYER, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 922-939.
- ROSENBAUM, D. A. (1980). Human motor initiation: Specification of arm direction and extent. *Journal of Experimental Psychology: General*, **109**, 444-474.