ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Memory and Language 49 (2003) 119–132

Journal of Memory and Language

www.elsevier.com/locate/jml

Effects of orthography on speech production in a form-preparation paradigm

Markus F. Damian* and Jeffrey S. Bowers

Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, UK

Received 9 October 2002; revision received 12 December 2002

Abstract

Four experiments investigated potential influences of spelling on single word speech production. A form-preparation paradigm that showed priming effects for words with initial form overlap was used to investigate whether words with form overlap, but different spelling (e.g., "camel"-"kidney") also show priming. Experiment 1 demonstrated that such words did not benefit from the form overlap, suggesting that the incongruent spelling disrupted the form-preparation effect. Experiment 2 replicated the first experiment with an independent set of items and an improved design, and once again showed a disruptive effect of spelling. To divert participants' attention from the spelling of the targets, Experiment 3 was conducted entirely in the auditory domain, but yielded the same outcome as before. Experiment 4 showed that matching initial letters alone, in the absence of matching sounds (e.g., "cycle"-"cobra"), did not produce priming. These findings raise the possibility that orthographic codes are mandatorily activated in speech production by literate speakers.

Language is supported by various subsystems, including phonology, orthography, semantics, and syntax. One of the most fundamental (and well studied) issues in psycholinguistics is the extent to which these subsystems act in a modular fashion. On the modular approach, outputs of one subsystem serves as input to another, with no feedback. For instance, according to the modular view of speech production developed by Garrett (1975), and Levelt and colleagues (e.g., Levelt, Roelofs, & Meyer, 1999), a non-linguistic conceptual system that encodes a speaker's thoughts activates the relevant lexical–semantic representations (semantic representations associated with syntax, often referred to as lemmas), which in turn retrieve lexical–phonological codes. Central to this approach is the claim that information processing is unidirectional, and that the retrieval of phonology in no way contributes, via feedback, to the selection of lemmas. On a non-modular approach, by contrast, the initial activation of lexical–semantic codes (based on conceptual inputs) leads to partial activation of phonology, which in turn feeds back and constrains the final lemma selection (Dell, 1986). On a stronger version of this approach, all the various subsystems would become activated in any language task, and interactions would occur between all systems. For instance, in speech production, not only would phonology feed back on semantics, but orthographic information would become activated as well and constrain the speech production process, despite the fact that orthographic information should be irrelevant to the process of speaking (for such a theoretical framework in the area of visual word perception, see, e.g., Van Orden & Goldinger, 1994).[1]

* Corresponding author. Fax: +44-117-928-8588.
E-mail address: m.damian@bristol.ac.uk (M.F. Damian).

[1] In multilinguals, mutual co-activation of components might even occur between languages; see Dijkstra, Timmermans, and Schriefers (2000).

0749-596X/03/$ - see front matter
doi:10.1016/S0749-596X(03)00008-1

Most work relevant to the issue of modularity has focused on interactions between processes that are integral to the task at hand. So in the case of speech production, semantics and phonology are two essential steps in the process of converting a thought to an utterance, and the most common question is whether phonology feeds back on semantics. Various findings have been taken as evidence that such feedback indeed occurs (e.g., Dell & O'Seaghda, 1992; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; but see, e.g., Levelt, 1992; Levelt et al., 1999, for a modular position). Similarly, in speech perception, the question is whether higher level information that must be accessed (e.g., complete phonological word forms, semantics, etc.) influences lower level perceptual processes (e.g., perception of phonemes). And again, various evidence suggests this is the case (e.g., Samuel, 1997), although there are modular interpretations of such findings as well (e.g., Norris, McQueen, & Cutler, 2000). In reading, the question is whether the identification of the visual structure of words (orthographic processing) is influenced by feedback from phonology or semantics, processes that are essential in the task of reading. Again, various evidence suggests that there is feedback from phonology (Stone, Vanhoy, & Van Orden, 1997), semantics (Hino & Lupker, 1996; Hino, Pexman, & Lupker, 2002; Pecher, 2001), and top-down processes within the orthographic system (McClelland & Rumelhart, 1981), although there are again different interpretations of these findings (cf. Grainger & Jacobs, 1994; Peereman, Content, & Bonin, 1998).

The current study considers potential interactions between language subsystems when some are not strictly relevant; in particular, we consider the role of orthography in constraining speech production. To the extent there are interactions, it would reflect a strong version of non-modularity, with diverse forms of linguistic knowledge automatically accessed (regardless of their relevance) and constraining speech production. There is in fact a small literature relevant to testing this strong version—most notably, the role of orthographic processes in speech perception. And as reviewed below, the findings often support a strong interactive position.

In a seminal study by Seidenberg and Tanenhaus (1979), participants either monitored lists of auditorily presented words for a target that rhymed with a cue word (Experiments 1 and 2), or performed rhyme judgements on auditorily presented pairs of words (Experiment 3). In both tasks, latencies were longer for cue–target pairs that rhymed but were orthographically dissimilar (rye-tie) than for those that were orthographically similar (pie-tie). Donnenwerth-Nolan, Tanenhaus, and Seidenberg (1981) replicated these findings while excluding the potential confound that orthographically similar cue–target pairs in Seidenberg and Tanenhaus' study might have simply been more predictable than dissimilar ones. Both studies suggest that orthographic information influences performance in tasks that do not require access to visual codes.

Taft and Hambly (1985) used a syllable monitoring task in which participants were asked to judge whether an auditorily presented syllable (such as "fin") occurred in an auditory target word (e.g., "finale"). Of special interest were target words in which the key syllable was on the unstressed syllable, hence the vowel was reduced ("metallic"; /mətælɪk/). When syllables were presented in which the vowel of the syllable was of full value (e.g., /mɛt/), participants tended to erroneously judge them as matching with the target word. This could be the case because the morphologically related word "metal" (/mɛtl/) exists, in which the vowel is full; possibly the auditory system automatically accesses the underlying common morphological code, and match/mismatch decisions are based on these codes. However, error rates were similarly high if syllable and target were not morphologically related (e.g., "lag", /læg/-"lagoon", /ləgun/). These results were attributed to an underlying orthographic code that was mandatorily activated in auditory processing. This code matched in both cases between syllable and target ("met"-"metallic"; "lag"-"lagoon") and obscured the phonetic mismatch between the full and the reduced vowel. Hence, spelling appeared to act as a powerful cue that induced participants to judge syllables and words as matching if their spelling also matched.

Findings reported by Dijkstra, Roelofs, and Fieuws (1995) also appear to support the claim that orthographic codes are involved in speech perception. In a phoneme monitoring task performed on Dutch words, response latencies were affected by the spelling of the targets: when target phonemes had more than one possible spelling (e.g., /k/ → "k" or "c"), responses were faster when they were presented in their primary spelling (e.g., /k/ in "kabouter") than when presented in their secondary spelling (e.g., /k/ in "cabaret"). These findings suggest that in addition to phonemic and phonological codes, orthographic codes also exert an influence in the phoneme monitoring task. Note, however, that Cutler, Treiman, and van Ooijen (1998) attributed this finding to participant strategies rather than automatic access to orthographic codes: they demonstrated that the effect disappeared when a large number of consistently spelled filler items was inserted that distracted participants' attention from the spelling of the stimuli.

The effects of orthography have also been investigated in *priming* paradigms: Jakimik, Cole, and Rudnicky (1985) asked participants to perform auditory lexical decisions to monosyllabic targets (e.g., "mess") that were preceded by multisyllabic auditory prime words (e.g., "message"). Facilitation was obtained only when spelling *and* sound of prime and target coincided (as in the example above), but not when they were related by sound alone ("definite"-"deaf") or by spelling alone ("legislate"-"leg"). Likewise effects were found for non-word targets ("regular"-"reg"). These results suggest that spelling facilitates processing in an auditory task, much like phonology facilitates processing in a visual task (Dijkstra, Frauenfelder, & Schreuder, 1993). On a related note, Frost, Repp, and Katz (1988) showed that when words masked with amplitude-modulated noise are presented simultaneously with printed words, processing of the auditory stimulus is facilitated if both printed and spoken words are congruent.

Ziegler and Ferrand (1998) showed that in an auditory lexical decision task, words with phonological rimes that could be spelled in multiple ways (e.g, /ip/ → "-eep" or "-eap") showed longer latencies and more errors than words whose rimes were uniquely spelled (e.g., /ʌk/ → "-uck"). This parallels previous findings obtained in visual lexical decisions with words that have many spellings given their sound, a so-called "feedback inconsistency" effect (e.g., Ziegler, Montant, & Jacobs, 1997). These effects were interpreted as indicating an interaction between visual and auditory codes evoked in both reading and listening tasks.

Hallé, Chéreau, and Segui (2000) utilized the fact that in French words such as "absurde," voice assimilation causes the phonetic realization of the first consonant to be /p/, not /b/ (/apsyrd/). In a gating experiment, participants indeed reported hearing the phonetically correct /p/. Yet, when they were presented with the entire word and asked to monitor for either /p/ or /b/, they exhibited a strong tendency to report /b/, which coincides with the spelling of "absurde," rather than /p/. These findings were taken to suggest a robust influence of orthographic codes onto phonetic perception.[2]

A related question is whether such effects might also be obtained in speech *production*, although the number of empirical studies that have explicitly investigated this question are fewer still. Tanenhaus, Flanigan, and Seidenberg (1980) conducted a study that used a modified form of the Stroop task (e.g., Warren, 1974): participants are asked to name the color of a target word which is preceded by a briefly presented prime word. The target itself is not a color word (as in the classic Stroop paradigm), but it still interferes to some extent with the color naming task, presumably because its lexical code is unintentionally activated. The activation level of the word can be manipulated by means of various relationships between the prime and the target word, such as whether or not they are associatively related. In Tanenhaus et al.'s study, visually or auditorily presented prime words could be either unrelated (well-food), phonologically and orthographically similar (mood-food), orthographically similar (good-food), or phonologically similar (rude-food) to the target. It was shown that both with auditory and visual primes, all three related conditions produced significant interference relative to the unrelated baseline. These results were interpreted in line with Seidenberg and Tanenhaus (1979) as indicating that both phonological and orthographic codes are automatically activated in both auditory and visual word recognition, and crucially influence speech production.

Lupker (1982) conducted a study in which participants were asked to perform timed naming responses to target pictures while attempting to ignore visual letter strings superimposed on the pictures. Crucially, relative to a control condition in which pictures and words were not form-related, distractors that were pronounced differently from the picture name, but shared spelling, produced substantial facilitation (picture: foot; distractor: boot; Experiment 1). Furthermore, distractors that shared the sound, but not the spelling with the picture label (picture: plane; distractor: brain) produced significantly less priming than distractors that shared both spelling and sound (picture: plane; distractor: cane; Experiment 2). These findings show that orthographic similarity between picture labels and distractor words affected picture naming latencies, in the absence of phonological similarity. Possibly, this implies that the naming of a picture evokes orthographic codes, which do or do not coincide with the orthographic information provided by the distractor word.

From a different perspective, Wheeldon and Monsell (1992) investigated long-term priming in speech production. In their first experiment, the naming of a picture (e.g., "shed") was preceded at various lags with a definition that elicited the picture label (e.g., "Building in which horses are kept"). Relative to an unprimed condition, substantial repetition priming was obtained. Experiments 2 and 3 investigated whether a parallel effect could be obtained with definitions that elicited *homophones* of the picture labels. Homographic homophones ("pipe"; as in a tool used for smoking or as used in plumbing) showed substantial priming whereas heterographic homophones ("sail"-"sale") did not. This

[2] On a related note, Morais, Cary, Alegria, and Bertelson (1979) showed that illiterate adults performed very poorly on phoneme addition and deletion tasks (such as to delete "p" from "purso," or to add "p" to "urso"), compared to matched controls, and Morais, Bertelson, Cary, and Alegria (1986) suggested that this poor performance was remedied in ex-illiterates, who performed similar to literate persons. Furthermore, Read, Zangh, Nie, and Ding (1986) demonstrated that Chinese non-alphabetic readers also lack phonemic awareness. However, as Morais and Kolinsky (1994) point out, these findings are probably attributable to a failure at a metaphonological level: phonemes are not consciously available to illiterates and non-alphabetic readers, but that does not imply that they do not constitute units of speech segmentation. In fact, it is well known that infants are rather apt at making phonemic distinctions (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971).

finding indicates that the spelling of the word pairs had a substantial impact on whether or not long-term priming was obtained, even though the orthographic codes of the homophones were never presented in the experiment. Once again, this might mean that even in speech production tasks, orthographic codes are automatically activated and contribute to lexical retrieval.[3]

As far as we are aware, these are the main studies that have assessed the potential effects of orthography on speech production. The research presented here attempts to contribute to this small literature by using a form-preparation task. This so-called *implicit priming* paradigm has in recent years become a prominent tool for investigating speech production (Chen, Chen, & Dell, 2002; Janssen, Roelofs, & Levelt, 2002; Meyer, 1990, 1991; Roelofs, 1996, 1998, 2002; Roelofs & Baayen, 2002; Roelofs & Meyer, 1998; see Levelt et al., 1999, for a review, and Roelofs, 1997, for a detailed computational account). Speakers in such an experiment first have to memorize small sets of word pairs such as *fruit-melon*, *iron-metal*, and *grass-meadow*. Then, during the following experimental block, their task is to produce the second word of each pair (response) in response to the visually presented first word (prompt). Prompts are repeatedly presented in random order and response times are measured relative to prompt onset. The crucial manipulation is whether or not all response words within a block overlap regarding certain phonological features. Blocks in which all response words share these characteristics are termed *homogeneous* while blocks in which the response words do not share these characteristics are are termed *heterogeneous*. Across all experimental blocks, each response word is produced an equal number of times, and merely the context in which it occurs is manipulated. The basic findings are that phonological overlap at the beginning of the response words yields a response time benefit whereas overlap in non-initial position does not, and that the benefit increases with increasing overlap (Meyer, 1990, 1991). Furthermore, the effect appears to be sensitive to abstract lexical properties such as the number of syllables and metrical patterns (see Roelofs & Meyer, 1998), which also suggests that it is based on phonological rather than on articulatory or memory processes. A computational account of this effect can be found in Roelofs (1997); it crucially hinges on the notion of partial planning. In homogeneous blocks, partial information about possible responses is present, and it is assumed that the speech production system assembles the response to the degree possible, resulting in a benefit over the heterogeneous condition in which no such preplanning is possible.

The following four experiments assessed potential effects of spelling in this paradigm. The study exploits the fact that the irregular orthography-to-phonology mappings of English allows for a number of words that share word-initial segments but differ in their spelling (e.g., "coffee"-"kennel"). If speech production is relatively encapsulated from other subsystems of language such as orthography, then spelling of the response words should be irrelevant—after all, participants never visually process the response words during the experimental blocks (although of course the word pairs are presented on the computer screen in the training phase prior to each block). On the other hand, if the mental lexicon is unified in a way that activation of one type of code results in the co-activation of corresponding codes in other subsystems, then retrieval of the response words could also activate their visual forms. In this case, it is possible that the difference in spelling might disrupt a benefit derived from the shared phonological segments.

Experiment 1

Method

Participants

Twenty-four undergraduate students at the University of Bristol participated in this experiment for course credit. All were native speakers and had normal or corrected-to-normal vision.

Materials and design

The independent variable in this experiment was type of context, with the three levels: homogeneous (all responses share the initial sound and spelling), heterogeneous (responses share neither initial sound nor spelling), and inconsistent (responses share initial sound, but differ in spelling). Materials were obtained from the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). All prompts and response words were nouns. The two segments /k/ and /ʤ/, whose English spelling is ambiguous, were employed. For each of the two segments, six prompt–response pairs with bi-syllabic response words were selected; three of the responses were spelled with either one of the possible two spellings in the initial position (e.g., "coffee" vs. "kennel"; see Appendix A). As a result, four sets of three items shared the initial segment as well as the spelling (*homogeneous* sets). Each of the six items within a set had a different vowel following the onset. The mean frequency of occurrence for the 12 response words was 8 per million in the CELEX database.

[3] Also see Perin (1983), showing that level of spelling abilities influenced the ease with which 14- and 15-year-olds could produce spoonerisms of the names of celebrities (e.g., "Phil Collins": /f/ ↔ /k/).

To construct the *heterogeneous* and the *inconsistent* sets, one item per triplet was swapped with one item from a different set. For instance, the triplet "camel," "coffee," "cushion" became "camel," "gypsy," "cushion" to form a set for the heterogeneous condition, and "kennel," "coffee," "cushion" to form a set for the inconsistent condition.[4] Semantic or associative relations between responses occurring together in a set were avoided. Across all 12 sets, each item was selected exactly once; hence, the same prompt–response word pairs were tested in all three conditions, and only the context in which they occurred was manipulated (see Appendix A for a complete listing of all sets).

Each participant was tested on each of the 12 blocks. The order in which the sets were presented was rotated across participants as follows: sets were presented in "superblocks" such that four consecutive sets were always of the same type of context (such as "homogeneous"). The order in which participants received these superblocks were determined by two sequentially balanced Latin Squares of size three, such that each superblock occurred in each position an equal number of times, and was followed and preceded by the other two types of context an equal number of times. Furthermore, the order in which each set was presented within each superblock was determined by a sequentially balanced Latin Square of size four. In this manner, each set occurred in each of the 12 positions equally often. The method of arranging and presenting "superblocks" was adapted from previous studies using this paradigm (e.g., Roelofs, 1997; Roelofs & Meyer, 1998).

In each experiment, the order in which items were presented within each block was random, with the constraint that immediate repetitions of pairs were excluded. A new random sequence was generated for each block and participant. Each block consisted of seven repetitions of each of the three stimuli, producing a total of 21 trials. Hence, the entire experiment consisted of 252 trials for each participant.

Apparatus

Stimuli were presented from an IBM-compatible computer on a 15 in. computer monitor using NESU (Nijmegen Experimental Set Up). The prompt words were presented in Helvetica 18-point bold font in uppercase letters in the center of the screen. Vocal responses were recorded with a microphone connected to the computer, which recorded a participant's response times to the nearest millisecond by means of a voice-activated relay.

Procedure

Participants were tested individually. They were informed that the experiment consisted of a series of relatively short experimental blocks; their task was to memorize a small set of word pairs presented on the screen prior to the beginning of each block, and to verbally produce the response word when cued by the prompt word during the experimental block. Participants were informed that they would have as much time as necessary to memorize the word pairs for each block. In the following, two practice blocks were carried out with items different from those used in the main experiment. Then, the 12 experimental blocks of 21 trials each were administered. Short breaks were provided in between blocks. Each experimental block was started by the experimenter as soon as participants indicated that they felt comfortable with the stimulus pairs for that block. The entire testing session lasted approximately 30 min.

On each trial, participants first viewed a fixation cross presented at the center of the screen for 500 ms. After a blank interval of 500 ms, the prompt word appeared for 1500 ms. Each naming response was judged by the experimenter to be either correct or incorrect (which included responses other than the expected ones, repairs, stuttering or mouth clicks, or malfunctioning of the voice key). Each trial was followed by a 1000 ms intertrial interval.

Results

Response times longer than 1500 ms or shorter than 200 ms (1.4%), as well as trials on which participants had made an error (2.9%) were excluded from the analysis. Furthermore, because context effects are being targeted and in order to reduce noise, latencies from the first occurrence of each stimulus within each block were also eliminated.

Table 1 presents the mean response latencies and error proportions for each of the three experimental conditions; each mean is based on 1728 observations (minus outliers and errors). A preparation effect was obtained only for the homogeneous, but not for the inconsistent, condition relative to the heterogeneous baseline.

Analyses of Variance (ANOVAs) that included type of context (heterogeneous vs. homogeneous vs. inconsistent) as a within-subjects or within-items variable were conducted on the mean response latencies. These yielded a highly significant outcome, $F_1(2,46) = 5.89$, $MSE = 6357$, $p = .005$; $F_2(2,22) = 8.37$, $MSE = 2978$, $p = .002$. Planned comparisons between the three conditions showed that the 28 ms difference between the heterogeneous and the homogeneous conditions was

[4] This so-called "odd-man-out" version of the implicit priming paradigm, in which only one out of several responses is not congruent with the others, is also used in Roelofs (1999) and Roelofs and Baayen (2002). We chose this manipulation because of the difficulties of finding inconsistent stimulus sets.

Table 1
Experiments 1–4: Mean response latencies (RT, in ms) and mean error proportions (PE, in %; standard deviations in brackets), varied by context (heterogeneous, homogeneous, inconsistent)

	RT	Effect	PE	Effect
Experiment 1				
Heterogeneous	653 (63)		2.9 (2.3)	
Homogeneous	625 (78)	28**	3.2 (2.1)	−0.3
Inconsistent	653 (74)	0	2.7 (2.4)	0.2
Experiment 2				
Heterogeneous	687 (92)		2.8 (2.1)	
Homogeneous	655 (91)	32**	2.5 (1.4)	0.3
Inconsistent	681 (98)	6	2.3 (1.4)	0.5
Experiment 3				
Heterogeneous	696 (110)		2.0 (2.8)	
Homogeneous	671 (114)	25*	2.2 (1.7)	−0.2
Inconsistent	694 (139)	2	2.7 (2.9)	−0.7
Experiment 4				
Heterogeneous	709 (80)		2.1 (1.6)	
Homogeneous	679 (93)	30*	1.9 (1.8)	0.2
Inconsistent	710 (77)	−1	2.0 (1.6)	0.1

* $p < .05$.
** $p < .01$.

significant, $t_1(23) = 2.80$, $p = .010$; $t_2(11) = 4.19$, $p = .002$. In contrast, the heterogeneous and the inconsistent conditions did not differ significantly from each other, $t_1(23) = 0.03$, $p = .977$; $t_2(11) = 0.10$, $p = .923$. Finally, the homogeneous and the inconsistent conditions differed significantly from each other, $t_1(23) = 2.87$, $p = .009$; $t_2(11) = 3.34$, $p = .007$, indicating a robust effect of spelling inconsistency.

ANOVAs performed on the error percentages yielded no significant effect of type of context, $F_1(2, 46) = 0.53$, $p = .590$; $F_2(2, 22) = 0.52$, $p = .600$. Likewise, planned tests that compared the errors in the three conditions with each other were not significant, $t_1(23) \leqslant 1.09$, $p \geqslant .288$; $t_2(11) \leqslant 1.08$, $p \geqslant .303$.

Discussion

The results from this experiment exhibit the robust form priming in the homogeneous relative to the heterogeneous condition previously reported (e.g., Meyer, 1990, 1991). Importantly, however, sets in which response words share their initial segment, but differ in their spelling, do not receive any benefit. This finding suggests an effect of orthographic codes in speech production. Possibly, retrieval of the phonological codes of the response words co-activates orthographic representations, which are incongruent and hence disrupt the phonological priming effect.

Such a mandatory co-activation of spelling information in speaking is not accounted for by any contemporary model of speech production. However, before the implications of these results for theories of language production are discussed, it is appropriate to consider alternative explanations for why the form priming effect disappears when words differ in their initial spelling. Remember that Cutler et al. (1998) demonstrated a strategic reliance on spelling codes in a phoneme monitoring task. Could the findings come about as a result of strategies that are specific to and evoked by the experimental task? The following two experiments investigate this possibility.

Experiment 2 has the following aims: first, it attempts to replicate the crucial result that inconsistent spelling of the initial letter disrupt the preparation effect with an independent set of items. Second, the experiment described above adopted its design from the various articles previously published on this paradigm in that the experimental conditions were presented in a blocked fashion (see, in particular, Roelofs, 1999). That is, for each participant, four successive blocks always stemmed from a particular experimental condition, and across participants, the order of these "superblocks" was varied by a Latin Square design. This design is problematic because the repeated succession of blocks from one and the same condition might have alerted participants to the manipulation. Experiment 2 improves on this design by alternating successive blocks from each condition such that two adjacent blocks never stem from the same condition. Finally, in post-experimental interviews, participants were asked about their guesses as to what was being investigated in the experiment; these interviews potentially reveal strategic reliance on spelling codes.

Experiment 2

Methods

Participants

Twenty-four undergraduate students at the University of Bristol, none of whom had been in Experiment 1, participated in this experiment for course credit. All were native speakers of English and had normal or corrected-to-normal vision.

Materials, design, and apparatus

A new set of 12 bi-syllabic nouns was selected from the CELEX database. Ideally, it would be desirable to employ different segmental contrasts from those used in the experiments above; however, word-initial spelling inconsistencies in the English language are not frequent, and adequate stimuli are hence difficult to find. Therefore, the two segments /k/ and /ʤ/ again served as the initial segments. Due to the serious constraints on stimulus selection, the stimulus "giant" used in Experiment 1 was also used here. The mean frequency of occurrence for the 12 response words was 5 per million in the CELEX database. Heterogeneous and inconsistent sets were constructed according to the procedure described in Experiment 1.

Again, each participant was tested on each of the 12 blocks. However, the order in which the sets were presented was now determined as follows: Conditions were rotated from block to block in a particular order, such as Block 1: "homogeneous," Block 2: "heterogeneous," Block 3: "inconsistent"; The order of the conditions was determined by two Latin Squares of size three such that four of the 24 participants received blocks in each order. Furthermore, the order of the four sets of each type of condition across the experiment was determined by a Latin Square of size four such that six participants received a particular order. In this way, each experimental block (such as, e.g., set 1—inconsistent) was presented an equal number of times in each of the 12 block positions.

The order in which items were presented within each block was random, with the constraint that immediate repetitions of pairs were excluded. A new random sequence was generated for each block and participant.

Design, apparatus, and procedure

These were identical to Experiment 1.

Results

The same procedure for data treatment as in the above experiment was applied, resulting in the exclusion of 1.3% of response time data points through absolute outliers, and 2.5% through errors. Table 1 presents the mean response latencies and error proportions for each of the three experimental conditions. Once again, a preparation effect was obtained only for the homogeneous, but not for the inconsistent, condition relative to the heterogeneous baseline.

ANOVAs conducted on the mean response latencies with the factor type of context (heterogeneous vs. homogeneous vs. inconsistent) as a within-subjects or within-items variable yielded a highly significant outcome, $F_1(2,46) = 11.66$, $MSE = 6,775$, $p < .001$; $F_2(2,22) = 15.90$, $MSE = 3462$, $p < .001$. Planned comparisons showed that the 32 ms difference between the heterogeneous and the homogeneous conditions was significant, $t_1(23) = 4.75$, $p < .001$; $t_2(11) = 7.58$, $p < .001$. In contrast, the heterogeneous and the inconsistent conditions did not differ significantly from each other, $t_1(23) = 0.80$, $p = .433$; $t_2(11) = 0.97$, $p = .353$. The homogeneous and the inconsistent conditions differed significantly from each other, $t_1(23) = 4.03$, $p < .001$; $t_2(11) = 4.19$, $p = .002$, indicating a robust effect of spelling inconsistency.

ANOVAs performed on the error percentages yielded no significant effect of type of context, $F_1(2,46) = 0.39$, $p = .678$; $F_2(2,22) = 0.56$, $p = .579$. Likewise, planned tests that compared the errors in the three conditions with each other were not significant, $t_1(23) \leqslant 0.85$, $p \geqslant .406$; $t_2(11) \leqslant 1.09$, $p \geqslant .301$.

Post-experimental interviews revealed that, not surprisingly, all participants had noticed that the blocking procedure manipulated the response-initial sound. However, none of them reported the spelling of the targets as a potential variable of interest, and virtually all of them expressed surprise when they were debriefed about the hypothesis underlying the experiment.

Discussion

The results provide a close replication to those from Experiment 1: once again, stimulus triplets that have overlapping word-initial phonemes, but inconsistent spelling did not show the priming effect arising from overlapping initial segments *and* spelling. This was the case with a new set of stimuli, despite the fact that the way in which the experimental blocks had been arranged made it less likely that participants' attention would be directed to the spelling of the words, and despite the fact that none of the participants reported awareness of the central hypothesis under investigation. Clearly, the spelling of the to-be-produced words exerts an influence on the latencies with which they are articulated.

One of the characteristics of the implicit priming paradigm, however, is that prompt–response pairs are visually presented in the memorization phase prior to each experimental block, which might alert participants to the orthographic characteristics of the response words. Furthermore, within the experimental blocks, prompt words are visually presented, which might additionally induce an orthographic processing mode in retrieving the response words. Separately or in combination, these task characteristics might induce participants to rely on visual codes; hence, the effects of spelling might not be representative of speech production in other tasks or in a natural context.

The following experiment specifically investigates the possibility that the effect of spelling might result from these experimental characteristics. It is identical in all relevant aspects to Experiment 1, except that it takes place entirely in the auditory domain: rather than being presented visually on the computer screen during the training phase, prompt–response pairs are repeatedly read aloud by the experimenter until participants indicate that they have memorized them. Within the experimental blocks, prompt words are presented auditorily via headphones. With no visual display throughout the experiment, participants should be discouraged from strategically adopting a processing mode that utilizes orthographic codes.

Experiment 3

Methods

Participants

Twenty-four undergraduate students at the University of Bristol, none of whom had been in the first two experiments, participated in this experiment for course credit. All were native speakers and had normal or corrected-to-normal vision.

Materials, design, and apparatus

These were identical to Experiment 1. However, prompt words were recorded by a male speaker and digitized with a sampling frequency of 16 kHz. Within the experimental blocks, prompts were presented to participants at a comfortable volume level over Sennheiser HD450 headphones.

Procedure

In contrast to Experiment 1, participants in Experiment 3 neither saw prompt words within the experimental blocks, nor prompt–response pairs in the training phase. In the training phase before each experimental block, the appropriate word pairs were repeatedly read out by the experimenter until the participant indicated that they were ready for the block. Within the experimental blocks, prompt words were presented auditorily.

Results

The same procedure for data treatment as in the above experiment was applied, resulting in the exclusion of 2.5% of response time data points through absolute outliers, and 2.3% through errors. Table 1 presents the mean response latencies and error proportions for each of the three experimental conditions. Once again, a preparation effect was obtained only for the homogeneous, but not for the inconsistent, condition relative to the heterogeneous baseline.

ANOVAs conducted on the mean response latencies with the factor type of context (heterogeneous vs. homogeneous vs. inconsistent) as a within-subjects or within-items variable yielded a marginally significant outcome in the analysis by subjects, $F_1(2,46) = 2.70$, $MSE = 4626$, $p = .078$, which was highly significant in the analysis by items, $F_2(2,22) = 9.36$, $MSE = 2377$, $p = .001$. Planned comparisons showed that the 25 ms difference between the heterogeneous and the homogeneous conditions was significant, $t_1(23) = 2.68, p = .014$; $t_2(11) = 4.02$, $p = .002$. In contrast, the heterogeneous and the inconsistent conditions did not differ significantly from each other, $t_1(23) = 0.15$, $p = .880$; $t_2(11) = 0.82$, $p = .432$. The homogeneous and the inconsistent conditions differed significantly from each other, $t_1(23) = 2.08$, $p = .049$; $t_2(11) = 3.47$, $p = .005$, indicating a robust effect of spelling inconsistency.

ANOVAs performed on the error percentages yielded no significant effect of type of context, $F_1(2,46) = 0.78$, $p = .465$; $F_2(2,22) = 1.09$, $p = .352$. Likewise, planned tests that compared the errors in the three conditions with each other were not significant, $t_1(23) \leqslant 1.48$, $p \geqslant .153$; $t_2(11) \leqslant 1.36$, $p \geqslant .202$.

Discussion

The results from Experiment 3 closely replicate those obtained in the first two experiments. This is the case despite the fact that the experiment was conducted entirely in the auditory domain, with no written material presented. Assuming that the auditory presentation format makes it less likely that participants directed their attention to the spelling of the targets, these findings suggest that access to orthographic codes is automatic in speech production.

Although we have been assuming that the preparation effect resides at the level of phonological encoding, it is important to note that the above findings are also consistent with the conclusion that the effect can be entirely attributed to an overlap in spelling. In most previous studies that have used this paradigm, spelling matched in the homogeneous condition (for a few exceptions see General discussion). Hence, the matching initial letters (and not the matching initial segments, as is normally assumed) might have supported the priming in the previous and present studies. The following experiment aims at isolating the effect of orthography by itself, in the absence of phonological overlap. It relies on the fact that in English, not only can a sound be spelled in multiple ways (which was manipulated in the above three experiments), but also that a letter can be pronounced in multiple ways. For instance, the words "cobra" and "cycle" share their initial spelling, but differ in their pronunciation. The final experiment exploits this fact by investigating whether orthography alone yields a preparation effect. Instead of using stimuli in the inconsistent condition which share the initial sound, but differ in their spelling, now we used stimuli that shared the spelling, but differed in their sound. This manipulation serves to further isolate the effects of spelling in speech production.

Experiment 4

Methods

Participants

Twenty-four undergraduate students at the University of Bristol, none of whom had been in the first three experiments, participated in this experiment for course credit. All were native speakers and had normal or corrected-to-normal vision.

Materials

A new set of 12 bi-syllabic nouns was selected from the CELEX database, which consisted of subsets of words that shared the same initial letter, but differed in their first phonological segment. The initial letters "c" (which can be pronounced /k/ or /s/) and "g" (which can be pronounced /g/ or /ʤ/) were used; for each letter, two sets of three items were selected such that the sets differed in their initial pronunciation. Each of the three items within a set had a different vowel following the onset. The results were four sets of items in which words shared both initial phoneme and spelling. The mean frequency of occurrence for the 12 response words was 21 per million in the CELEX database. To construct heterogeneous and inconsistent sets, one item per triplet was swapped with one item from another set. Hence, in heterogeneous sets, one out of three items differed in both initial segment and spelling (e.g., "cobra," "giant," "candle"). Crucially, in inconsistent sets, items shared the initial letter, but differed in their initial sound (e.g., "census," "climate," "candle"). Appendix C provides a full listing of all sets.

Design, apparatus, and procedure

These were identical to Experiment 1.

Results

The same procedure for data treatment as in Experiments 1, 2, and 3 was applied, resulting in the exclusion of 2.7% of response time data points through absolute outliers, and 2.0% through errors. Table 1 presents the mean response latencies and error proportions for each of the three experimental conditions. Once again, a preparation effect was obtained for the homogeneous condition relative to the heterogeneous baseline. In contrast, the inconsistent condition—identical initial spelling but different initial phoneme—showed no priming effect relative to the baseline.

ANOVAs conducted on the mean response latencies with the factor type of context (heterogeneous vs. homogeneous vs. inconsistent) as a within-subjects or within-items variable yielded a significant outcome, $F_1(2, 46) = 3.45$, $MSE = 7440$, $p = .040$; $F_2(2, 22) = 5.70$, $MSE = 3988$, $p = .010$. Planned comparisons showed that the 30 ms difference between the heterogeneous and the homogeneous conditions was significant, $t_1(23) = 2.30$, $p = .031$; $t_2(11) = 2.59$, $p = .025$. In contrast, the heterogeneous and the inconsistent conditions did not differ significantly from each other, $t_1(23) = 0.11$, $p = .912$; $t_2(11) = 0.08$, $p = .940$. The homogeneous and the inconsistent conditions differed significantly from each other, $t_1(23) = 2.12$, $p = .045$; $t_2(11) = 2.70$, $p = .021$.

ANOVAs performed on the error percentages yielded no significant effect of type of context, $F_1(2, 46) = 0.10$, $p = .906$; $F_2(2, 22) = 0.12$, $p = .887$. Likewise, planned tests that compared the errors in the three conditions with each other were not significant, $t_1(23) \leqslant 0.43$, $p \geqslant .763$; $t_2(11) \leqslant 0.43$, $p \geqslant .674$.

Discussion

The results are clear in showing that matching initial spelling, in the absence of an overlap in phonological characteristics, is insufficient to cause a response time benefit relative to an unrelated baseline. These findings suggest that the basic form-preparation effect cannot be attributed to processes taking place at the orthographic level. Instead, as previously assumed, it seems to be the case that the effect resides at the level of phonological encoding, with feedback from orthography constraining phonological encoding, as discussed below.

General discussion

Four experiments using a form-preparation paradigm demonstrated effects of orthography in speech production: the response time benefit deriving from a word-initial segmental overlap (originally demonstrated by Meyer, 1990, 1991, and replicated here) is disrupted if words within a block have the same initial segment, but this segment is spelled in more than one way. At the same time, overlapping spelling, but mismatching initial segments are not enough to generate the effect. The study thus provides evidence for interactions between orthography and phonology in speech production, consistent with the few studies that have observed similar results in perception (and even fewer in production). Accordingly, the results provide some support for a strong non-modular approach in which non-relevant linguistic knowledge can be activated and feed back on relevant systems in order to constrain language processing.

The assumption underlying the current article is that the priming effect in the implicit priming paradigm reflects partial phonological encoding in the homogeneous condition, and hence allows insights into the workings of the language production system. An issue that is common to all studies that use this paradigm (Chen et al., 2002; Janssen et al., 2002; Meyer, 1990, 1991; Roelofs, 1996, 1998, 1999, 2002; Roelofs & Baayen, 2002; Roelofs & Meyer, 1998; Santiago, 2000) is whether this assumption is correct, or whether the priming effect rather arises from different mechanisms, such as articulatory or memory processes. According to one possible alternative account, the shared word-initial sounds of homogeneous sets allow the speaker to move the articulators into the correct starting position, hence the time required to initiate a response will be shorter than that in heterogeneous sets. This hypothesis, centering on articulatory (rather than phonological) preparation, is difficult to reconcile with the finding that the priming effect grows in size with increasing segmental overlap, even into the second syllable (Meyer, 1990). And perhaps more important, effects of morpheme frequency (Roelofs, 1996) and other abstract linguistic variables (e.g., Roelofs, Meyer, & Levelt, 1998) are difficult to account for by an articulatory account.

A potentially more serious problem is that the priming effect can be attributed to memory retrieval processes. According to this account, when participants acquire the prompt–response pairings in the study phase, they rehearse them together and establish an episodic association between them. In the experimental blocks, they are presented with the prompt word as a retrieval cue, and generate the associated word by following the episodic association. Crucially, homogeneous and heterogeneous sets are not necessarily constant with regard to the number and effectiveness of retrieval cues: in homogeneous sets, shared orthographic and/or phonological fragments might constitute additional episodic retrieval cues that facilitate memory retrieval, relative to heterogeneous sets in which there are no such cues. Hence, the priming effect might be attributable not to partial planning in language production, but rather simply to the fact that targets in homogeneous sets are easier to retrieve from memory.

One form of evidence against this possibility comes from the finding that the priming effect shown in the implicit priming paradigm can also be obtained when the task is simply to name pictures (Roelofs, 1999, Experiment 3; Santiago, 2000). Naming pictures is clearly less dependent on episodic retrieval components than the prompt–response generation task used in the standard form-preparation paradigm, but both tasks share the word production component. The fact that the form-preparation benefit can be found in both tasks strongly argues against the possibility that memory retrieval cues are the crucial factor. Furthermore, as noted above, there is now a good amount of research that has demonstrated that the priming effect is sensitive to a whole host of rather abstract linguistic variables (such as suprasegmental structure; Roelofs et al., 1998). It is difficult to see how some of these variables could be reduced to retrieval cue effectiveness. In our own study, the similar pattern of priming obtained when the prompts were presented in visual or auditory format, and the finding that participants were unaware of the spelling manipulation also lends some support to the conclusion that strategic episodic effects did not play a role in our results.[5]

Although orthographic effects on speech production processes might seem surprising (perhaps even implausible), we would argue that this claim makes sense in the context of various results suggesting strong bi-directional interactions between the orthographic and the phonological subsystems. In the literature on visual word recognition this assumption is ubiquitous (for experimental evidence see, e.g., Pexman, Lupker, & Jared, 2001; Stone et al., 1997), and a number of computational models of reading have implemented bi-directional links between the orthographic and the phonological lexicon that would allow orthographic ef-

[5] In addition, there are rather specific findings obtained with this paradigm that suggest that the priming effect cannot be reduced to the effectiveness of sets of words as retrieval cues. For instance, Roelofs (1998) investigated the issue of morphological processes by means of particle–verb combinations (e.g., "look up") as targets. A first experiment showed that sets that shared the initial particle, such as "*op*zoeken"-"*op*draaien"-"*op*geven" (look up, wind up, give up), yielded priming over sets with heterogeneous particles ("*op*zoeken," "*af*draaien," "*uit*geven"; look up, show, spend). In contrast, when the verb, rather than the particle, overlapped, no priming effect was found ("op*zoeken*," "af*zoeken*," "uit*zoeken*" vs. "op*zoeken*," "af*draaien*," "uit*geven*"). In a second experiment, the same particle–verb combinations were tested in a form in which their order was reversed, namely in their imperative form (such as "zoek op!," look up!). Here, the opposite pattern was found: shared verbs, but not shared particles, yielded a priming effect. These findings, according to Roelofs, refute ease of memory retrieval as the central variable, because particle and verb combinations were the same in both experiments, and merely their order differed. In contrast, an incremental model or phonological encoding that embodies left-to-right planning can accommodate these findings by assuming that utterance-initial overlap allows partial planning and hence results in a latency benefit.

fect to occur in auditory-oral language. For instance, the DRC model of word recognition and naming (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) implements such bi-directional links. Hence, not only does it allow transmission of activation along this "lexical" route, but it could also in principle allow feedback of the orthographic onto the phonological lexicon. A similar structure is implemented in Jacobs, Rey, Ziegler, and Grainger's (1998) MROM-p model. These models included bi-directional feedback in response to various experimental results (such as those cited above). As long as the same phonological system is involved in speech production and reading (which has not been questioned to our knowledge), then orthographic activation (and feedback to phonology) should be expected.

This is not to suggest that all forms of linguistic knowledge contribute to whatever linguistic task. For instance, Franck, Bowers, Frauenfelder, and Vigliocco (in press) assessed the role of orthography in computing subject–verb agreement when producing written and spoken sentences. Past research has shown that the phonological form of a word contributes to agreement in spoken production: for example, Italian speakers make fewer errors in inflecting a verb for number when the subject of a sentence is phonologically marked as singular or plural (e.g., il paese/i paesi; the village/the villages) than when it is unmarked (e.g., la citta'/le citta'; the town/the towns; Vigliocco, Butterworth, & Semenza, 1995). These finding were taken to indicate that phonological activation of a word feeds back on the syntactic processor to constrain agreement. Franck et al. found the parallel finding in written French production, with fewer agreement errors when writing verbs when the subject of sentences was orthographically marked for number (e.g., chanson/chansons, for song/songs) compared to unmarked (e.g., secours/secours, for rescue/rescues). Critical for present purposes, however, the study failed to observe any effect of orthography on subject–verb agreement for the same items in spoken production (that is, a similar number of errors occurred for orthographically marked and unmarked items), suggesting that form influences on syntax are restricted to a specific modality.

Any definite account as to how the initial spelling of the stimuli in the form-preparation task matters is of course speculative at this point. However, if one assumes that phonological segments are connected to corresponding letters or graphemes (e.g., the sublexical route of dual-route accounts of reading, but in reverse), then it could be that sounds and letters have to enter a stable state of congruency to be selected or planned (e.g., Stone et al., 1997; Ziegler & Ferrand, 1998). 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. No such conflict would occur in blocks in which stimuli are congruent in both sound and spelling.

In addition to general theoretical considerations, the finding that spelling influences performance on the form preparation paradigm highlights an important methodological issue. The numerous studies that have employed this technique have not considered the relevance of spelling, and accordingly, effects attributed to phonology may have another source. For example, this confound is a problem for a study by Roelofs (1999). The study addressed the issue whether phonological segments or featural representations constitute the basic units of speaking at the phonological level. Roelofs used the form-preparation paradigm to investigate whether entire word-initial phonological segments had to be shared to produce the priming effect, or whether a high degree of featural overlap was sufficient. Response words with a voiced word-initial consonant (e.g., /b/) were compared to those that started with their unvoiced counterpart (e.g., /p/). More specifically, in addition to homogeneous sets such as the Dutch words *been*, *bos*, *baard* and heterogeneous sets such as *been*, *dolk*, *film*, sets of high featural overlap were used, such as *bos*, *been*, *pet*. The results exhibited the response time benefit for the homogeneous over the heterogeneous sets previously reported, but no priming effect for the sets with high featural overlap, but mismatching segments. These results were taken to imply that entire segments are the planning units of phonological coding in speech production.

Of course, the response words in the high featural overlap condition not only have different word-initial segments, but also differ in their spelling. Roelofs' (1999) conclusion that high featural overlap does not benefit phonological encoding rests on the implicit assumption that the difference in spelling is irrelevant. The findings reported here, in contrast, suggest that incongruent spelling alone, apart from any featural mismatch, would have disrupted the priming effect. Hence, the question of whether segments or features are the basic units of phonological planning remains unresolved.

Before concluding, it is important to note one discrepancy between our findings and those of Meyer (1990, 1991) who reported five experiment in which one out of the three or four stimulus sets was inconsistent with regard to initial spelling (e.g., "sinas, citer, silo, sisal, sieraad" in Experiment 1, 1990). When these experiments were analyzed separately by set, the preparation effect was eliminated in one case (1991; Experiment 7)—consistent with our findings—but the other sets showed varying degrees of priming. Likewise, Roelofs et al. (1998) reported one experiment (Experiment 2) in which three out of 32 sets of stimulus triplets were of the inconsistent type, but as the authors did not list priming effects separately for each set, it is difficult to judge what effects inconsistent spelling had on the overall effects.

The instances of preserved priming reported by Meyer (1990, 1991) are obviously not congruent with those reported in the current study. However, we did replicate our Experiment 1 with a different population (students from Rice University, Houston) and with two different stimulus sets (Experiments 1 and 3), hence we are reasonably confident about the validity of our results. One possible, albeit admittedly ad hoc, explanation for this discrepancy is that whereas our experiments were conducted in English, Meyer's experiments were conducted in Dutch. Dutch, despite having occasional word-initial inconsistencies such as those appearing in Meyer's experiments, overall embodies rather regular orthography-to-phonology mappings. It is at least possible that, due to the high extent of inconsistency present in English, effects of orthography are more pronounced here than in other languages that are more regular, such as Dutch. But clearly, further work will need to be carried out in order to better understand the basis of these mixed results.

To return to the main question of interest, the results reported here merely constitute an initial attempt to investigate the potential influences of spelling in speech production. For a number of reasons, the paradigm chosen for the current article has limitations: first, only spelling inconsistencies at the *beginning* of a word can be employed. Only a few instances of such inconsistencies exist in English, which—together with other constraints on stimulus selection—render the possible number of testable contrasts rather small. It would clearly be advantageous to use tasks that were able to target spelling inconsistencies in the body of the word, such as the "cream"-"theme" word pairs used in Seidenberg and Tanenhaus' (1979) study. However, such tasks remain to be developed. Also, the basic effect in the implicit priming paradigm critically relies on the repeated production of a small number of stimuli. Despite our efforts in Experiments 2 and 3 to minimize the possibility that participants strategically directed their attention to the spelling of the stimuli, it remains possible that this paradigm does not adequately represent speech production in a more natural context. Clearly, the hypothesis that orthographic codes are mandatorily evoked in speaking should in future work be confirmed by converging evidence from a variety of experimental paradigms.

Appendix A. Stimuli used in Experiments 1 and 3

Context: Homogeneous
Set 1: camel, coffee, cushion
Set 2: kennel, kayak, kidney
Set 3: giant, gypsy, genius
Set 4: jasmine, jewel, joker

Context: Heterogeneous
Set 5: camel, gypsy, cushion
Set 6: kennel, jewel, kidney
Set 7: giant, coffee, genius
Set 8: jasmine, kayak, joker

Context: Inconsistent
Set 9: kennel, coffee, cushion
Set 10: camel, kayak, kidney
Set 11: jasmine, gypsy, genius
Set 12: giant, jewel, joker

Appendix B. Stimuli used in Experiment 2

Context: Homogeneous
Set 1: candle, cobra, climate
Set 2: kettle, kodak, kilo
Set 3: genie, giant, gymnast
Set 4: jester, jaguar, jury

Context: Heterogeneous
Set 5: candle, giant, climate
Set 6: kettle, jaguar, kilo
Set 7: genie, kodak, gymnast
Set 8: jester, cobra, jury

Context: Inconsistent
Set 9: kettle, cobra, climate
Set 10: candle, kodak, kilo
Set 11: jester, giant, gymnast
Set 12: genie, jaguar, jury

Appendix C. Stimuli used in Experiment 4

Context: Homogeneous
Set 1: cobra, climate, candle
Set 2: census, city, cycle
Set 3: genie, giant, gymnast
Set 4: garden, giggle, gutter

Context: Heterogeneous
Set 5: cobra, giant, candle
Set 6: census, giggle, cycle
Set 7: genie, climate, gymnast
Set 8: garden, city, gutter

Context: Inconsistent
Set 9: census, climate, candle
Set 10: cobra, city, cycle
Set 11: garden, giant, gymnast
Set 12: genie, giggle, gutter

References

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (Release 2) [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

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 and Language, 46*, 751–781.

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. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

Cutler, A., Treiman, R., & van Ooijen, B. (1998). Orthographik inkoncistensy ephekts in foneme detektion?. In *Proceedings of the fifth international conference on spoken language processing* (pp. 2783–2786). Sidney: ICSLP.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93*, 283–321.

Dell, G. S., & O'Seaghda, P. G. (1992). Stages of lexical access in language production. *Cognition, 42*, 287–314.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*, 801–838.

Dijkstra, T., Frauenfelder, U., & Schreuder, R. (1993). Bidirectional grapheme–phoneme activation in a bimodal detection task. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 931–950.

Dijkstra, T., Roelofs, A., & Fieuws, S. (1995). Orthographic effects on phoneme monitoring. *Canadian Journal of Experimental Psychology, 49*, 264–271.

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language, 42*, 445–464.

Donnenwerth-Nolan, S., Tanenhaus, M. K., & Seidenberg, M. S. (1981). Multiple code activation in word recognition: Evidence from rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 170–180.

Eimas, P., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303–306.

Franck, J., Bowers, J., Frauenfelder, U., & Vigliocco, G. (in press). Orthographic influences on agreement: A case for modality-specific form effects on grammatical encoding. *Language and Cognitive Processes*.

Frost, R., Repp, B. H., & Katz, L. (1988). Can speech perception be influenced by simultaneous presentation of print? *Journal of Memory and Language, 27*, 741–755.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 9, pp. 133–177). New York: Academic Press.

Grainger, J., & Jacobs, A. M. (1994). A dual read-out model of word context effects in letter perception: Further investigations of the word superiority effect. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1158–1177.

Hallé, P. A., Chéreau, C., & Segui, J. (2000). Where is the /b/ in "absurde" [apsyrd]? It is in French listeners' minds. *Journal of Memory and Language, 43*, 618–639.

Hino, Y., & Lupker, S. J. (1996). The effects of polysemy in lexical decision and naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1331–1356.

Hino, Y., Pexman, P. M., & Lupker, S. J. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 686–713.

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

Jakimik, J., Cole, R. A., & Rudnicky, A. I. (1985). Sound and spelling in spoken word recognition. *Journal of Memory and Language, 24*, 165–178.

Janssen, D. P., Roelofs, A., & Levelt, W. J. M. (2002). Inflectional frames in language production. *Language and Cognitive Processes, 17*, 209–236.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition, 42*, 1–22.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–75.

Lupker, S. J. (1982). The role of phonetic and orthographic similarity in picture–word interference. *Canadian Journal of Psychology, 36*, 349–376.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375–407.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables. *Journal of Memory and Language, 29*, 524–545.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30*, 69–89.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition, 24*, 45–64.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*, 323–331.

Morais, J., & Kolinsky, R. (1994). Perception and awareness in phonological processing—The case of the phoneme. *Cognition, 50*, 287–297.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299–325.

Pecher, D. (2001). Perception is a two-way junction: Feedback semantics in word recognition. *Psychonomic Bulletin & Review, 8*, 545–551.

Peereman, R., Content, A., & Bonin, P. (1998). Is perception a two-way street? The case of feedback consistency in visual word recognition. *Journal of Memory and Language, 39*, 151–174.

Perin, D. (1983). Phonemic segmentation and spelling. *British Journal of Psychology, 74*, 129–144.

Pexman, P. M., Lupker, S. J., & Jared, D. (2001). The nature of homophone effects and phonological processing in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 139–156.

Read, C., Zangh, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition, 24*, 31–44.

Roelofs, A. (1996). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language, 35*, 854–876.

Roelofs, A. (1997). 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 spech production: Verb–particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 904–921.

Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes, 14*, 173–200.

Roelofs, A. (2002). Spoken language planning and the initiation of articulation. *The Quarterly Journal of Experimental Psychology A, 55*, 465–483.

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, and Cognition, 24*, 922–939.

Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1998). A case for the lemma/lexeme distinction in models of speaking: Comment on Caramazza and Miozzo (1997). *Cognition, 69*, 219–230.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology, 32*, 97–127.

Santiago, J. (2000). Implicit priming of picture naming: A theoretical and methodological note on the implicit priming task. *Psicológica, 21*, 39–59.

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 546–554.

Stone, G. O., Vanhoy, M. D., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language, 36*, 337–359.

Taft, M., & Hambly, G. (1985). The influence of orthography on phonological representations in the lexicon. *Journal of Memory and Language, 24*, 320–335.

Tanenhaus, M. K., Flanigan, H. P., & Seidenberg, M. S. (1980). Orthography and phonological activation in auditory and visual word recognition. *Memory & Cognition, 8*, 513–520.

Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance, 1994*, 1269–1291.

Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Computing Subject Verb agreement in speech: The role of semantic and morphological information. *Journal of Memory and Language, 34*, 186–215.

Warren, R. E. (1974). Association, directionality, and stimulus encoding. *Journal of Experimental Psychology, 70*, 102–108.

Wheeldon, L., & Monsell, S. (1992). The locus of repetition priming of spoken word production. *Quarterly Journal of Experimental Psychology A, 44*, 723–761.

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review, 5*, 683–689.

Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language, 37*, 533–554.