Journal of Experimental Psychology:
Learning, Memory, and Cognition
2005, Vol. 31, No. 5, 921–932

0278-7393/05/$12.00 DOI: 10.1037/0278-7393.31.5.921

Semantic and Phonological Context Effects in Speech Error Repair

Robert J. Hartsuiker
Ghent University

Martin J. Pickering
University of Edinburgh

Nivja H. de Jong
University of Amsterdam

When speakers repair speech errors, they plan the repair in the context of an abandoned word (the error) that is usually similar in meaning or form. Two picture-naming experiments tested whether the error's lexical representations influence repair planning. Context pictures were sometimes replaced with target pictures; the picture names were related in meaning or form or were unrelated. The authors measured target picture-naming latencies separately for trials in which the context name was interrupted or completed. Interrupted trials showed semantic interference and phonological facilitation, whereas completed trials showed semantic facilitation and phonological interference. Thus, errors influence repair production. The authors explain the polarity of these effects in terms of the literature on context effects in word production.

Keywords: verbal self-monitoring, speech error repair, word production, picture naming

We sometimes do not notice our speech errors, but on other occasions we detect the error, interrupt ourselves, and repair the mistake; that is, we replace the words with the ones we intended to say. These processes are illustrated in Examples 1–2 (errors in italics).

> Example 1: You cannot even get a job in an English hospital without passing *an English* a French test. (Blackmer & Mitton, 1991)

> Example 2: Okay, well go between the *gold line* gold mine and the outlaws' hideout. (Edinburgh, Scotland, Map Task Corpus; Anderson et al., 1991)

The processes of error detection, interruption, and repair are referred to as *self-monitoring* (Hartsuiker & Kolk, 2001; Levelt, 1989; Postma, 2000). Whereas most research on self-monitoring has considered the question of how we detect errors, few studies have considered how we repair. However, having to repair an error presents the speaker with a situation that is importantly different from error-free word production. Because most speech errors are either related in meaning (see Example 1) or in form (see Example 2) to the intended word (Fromkin, 1971; Garrett, 1975; Harley, 1984), the production of the repair takes place in the context of just having said a very similar word. It is conceivable, therefore, that lexical and sublexical representations corresponding to the error are still active while the repair is being produced. This poses an interesting problem: Do these active representations hinder the planning of the repair (e.g., because they are likely to be selected again) or do they help the planning of the repair (e.g., because they prime the correct representation)?

The goal of this study is to test whether word production is affected by a meaning or form relation with a word that was just abandoned, and if so, whether these relations facilitate or interfere with production. To study these processes experimentally, we used a paradigm in which speakers name one picture (the context picture), which sometimes changed in another picture (the target picture), and we varied whether the picture names were related or not. This situation is, of course, different from cases in which people realize they are describing something incorrectly and is more similar to cases in which they realize they have misperceived the object that they are describing. However, similar to the situation of speech error repair, in this paradigm an abandoned word may still be active and influence the subsequent response.

Studies on speech errors have shown that although some errors are completely unrelated to the intended word, most of them are related to the intended word in meaning, form, or both (see Dell, Schwartz, Martin, Saffran, & Gagnon, 1997). A number of theorists explicitly assume error-to-repair priming (Hartsuiker & Kolk, 2001; Postma & Kolk, 1993) in these related cases. For example, Postma and Kolk (1993) suggested that residual activation of a phonological error helps the production of the repair, because the overlapping phonological representations are primed. They applied their theory to stuttering and argued that priming of partially correct phonology, as a result of repeated repair attempts, eventually helps stuttering speakers overcome difficulties with retrieving

Robert J. Hartsuiker, Department of Experimental Psychology, Ghent University; Martin J. Pickering, Department of Psychology, University of Edinburgh; Nivja H. de Jong, Amsterdam Center for Language and Communication, University of Amsterdam.

This study was supported by a grant from the Economic and Social Research Council (R000223812) awarded to Robert J. Hartsuiker and Martin J. Pickering. We thank Ciara Catchpole and Manon Jones for their help with running the experiments and measuring the response times, Keith Laws for making his visual overlap data available to us, and Jan Lammertyn for implementing a program to determine visual overlap.

Correspondence concerning this article should be addressed to Robert J. Hartsuiker, Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000, Ghent, Belgium. E-mail: robert.hartsuiker@ugent.be

a word's phonology. However, note that it is also possible that residual activation of phonological representations hinders the production of the repair because of competition between the phonological representations that do not overlap. A well-known example of production problems associated with phonological similarity is the severely disruptive effect of phoneme repetitions in tongue-twisters such as "she sells sea shells by the sea shore" (e.g., Postma & Noordanus, 1996; Wilshire, 1998, 1999). Similarly, semantic errors share certain semantic characteristics with their repairs, but they correspond to different words, and these words may compete for selection. Thus, for semantic repairs, too, it remains to be seen whether activation of the error helps or hinders.

In contrast to the view that the lexical representations of abandoned words (such as errors) remain active and can therefore influence the production of subsequent words, it is also possible that speakers exert certain control mechanisms that prevent the abandoned word from being selected again. For example, Levelt (1989) proposed that interruption entails that a signal to stop is sent to all major components of the language production system (in Levelt's terms, the *conceptualizer*, the *formulator*, and the *articulator*). It is not clear how such a signal would affect the representations at each level, but in one interpretation, there would be sustained inhibition, which would "wipe clean" the activation of incorrect representations currently being produced. If so, one would expect no effect of relatedness on repair production, because the abandoned word is prevented from becoming active and, hence, cannot influence the production of subsequent words.

It is important to note that the question of whether errors affect repairs may depend on whether speakers interrupt the production of the error. When speakers repair, they sometimes interrupt within the error (see Example 3), but at other times they say the error completely (see Example 4).

Example 3: Straw— . . . Grape

Example 4: Strawberry . . . Grape

Analyses of corpora of naturalistic or experimentally elicited repairs have revealed that incidents such as that in Example 3 constitute about 20% of error repairs (Berg, 1986; Brédart, 1991; Levelt, 1983). There are two reasons why errors may affect repairs differently, depending on whether the error is interrupted. First, according to some authors (e.g., Berg, 1986) there is a qualitative difference between repairs of interrupted and uninterrupted errors. This contention is supported by the finding that for so-called "appropriateness repairs," the error is almost never interrupted. These are incidents in which the problematic word provided information in an inappropriate way, for example at the wrong level of specificity for the context (e.g., saying "grape" instead of "white grape"). Berg therefore argued for a processing component that decides whether to interrupt and, if so, when to interrupt.

Second, there may also be a quantitative difference between interrupted and noninterrupted errors. In particular, Brédart (1991) demonstrated that the probability of interrupting a word depends on the length of that word in syllables. Word-internal interruptions almost never happened when the erroneous word was monosyllabic but were much more frequent with polysyllabic words. This suggests that the difference between incidents such as those in Examples 3 and 4 is a matter of time course: If the error is monosyllabic, the moment of error detection and the subsequent signal to interrupt are too late to prevent the error from slipping through. It is important that if error-detection and subsequent interruption occur relatively early for interrupted words, it is likely that the process of repair also starts relatively early. This follows from the assumption that the processes of interruption and repair are coupled in time (Hartsuiker & Kolk, 2001; Levelt, 1989). It is therefore conceivable that the error's lexical representations are more active and, therefore, affect the repair more when the error is interrupted.

Until now, we have talked about an error's lexical and sublexical representations. However, before we can derive predictions about the effect of residual activation of such representations on error repair, it is important to specify which representations we minimally need to postulate in word production. In this article, we largely follow the theory of word production proposed by Levelt, Roelofs, and Meyer (1999). We assume that the production of words requires the activation of (a) lexical concepts, (b) lemmas, (c) word forms, (d) phoneme units, and (e) word frame units. These units, and their connectivity, are illustrated in Figure 1.

Word production begins with the identification of a lexical concept to express, for example, as a result of perceiving an object (e.g., a strawberry) and being instructed to name that object. When the speaker intends to lexicalize a certain concept—for example,

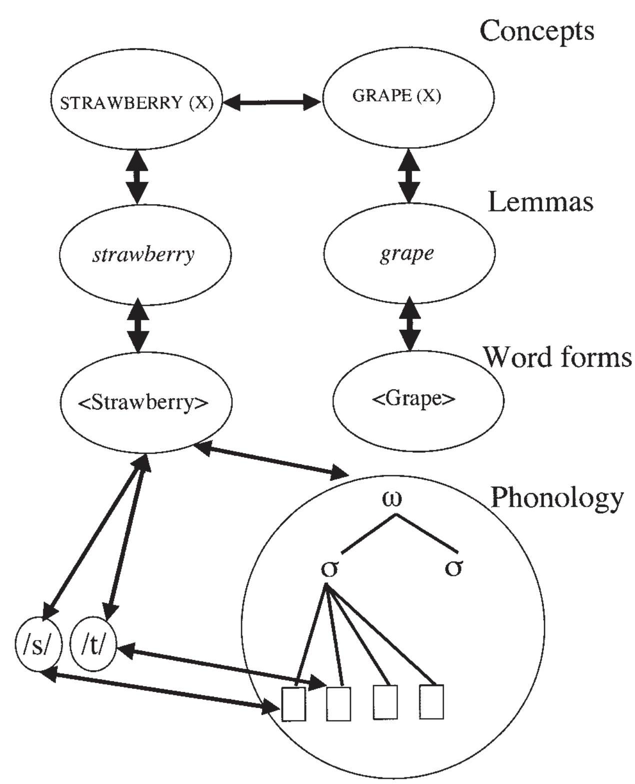

Figure 1. Schematic view of the minimally necessary representations we postulate in word production, with respect to the example words *strawberry* and *grape* (only concepts and lexical units displayed for the latter word). Following Dell (1986, 1988), we include bidirectional connections between each level, although for our account, only bidirectionalities between concepts and lemmas and between phonemes and word forms are essential.

STRAWBERRY(X)—activation spreads to its lexical representation (*strawberry*), and activation from the lexical level feeds back to the conceptual level (Levelt et al., 1999). Throughout this article, we assume the distinction between lemmas and word forms for expository purposes, but we note that this distinction has been challenged (Caramazza, 1997). We follow Levelt et al. (1999) and Dell (1988) in assuming that following lexical retrieval, people phonologically encode the word by first spelling out the components of the word. This includes units for phonemes (e.g., /s/, /t/) and its word frame (e.g., a frame specifying that there are two syllables, with stress on the first one and with slots for the phonemes of each syllable). In the next step, the retrieved phonemes are assigned to positions in the word frame, resulting in a phonological word (see Levelt & Wheeldon, 1994). On the basis of this phonological word, a speech motor code is retrieved.

We hypothesized that the production of an error results in residual activation for each of these representations (at least for the situation in which the error is said completely). But how does this activation impact upon repair production? We are not aware of any studies that have empirically tested for effects of errors on repair production, but it has been a common research strategy to investigate context effects on error-free production. In particular, a large number of studies have tested for effects of a semantically, phonologically, or unrelated distracter word on the production of a target word (e.g., Bloem & La Heij, 2003; Cutting & Ferreira, 1999; Damian & Martin, 1999; Glaser & Glaser, 1989; Lupker, 1979; A. S. Meyer & Schriefers, 1991; Rayner & Posnansky, 1978; Roelofs, 2003; Schriefers, Meyer, & Levelt, 1990; Wheeldon, 2003). In all these tasks, a word is produced (by reading, responding to a definition, naming a picture, or translating) in the context of a distracter word or picture. Thus, these studies have at least superficial similarities with the situation of speech-error repair.

Although these studies have compellingly demonstrated that a semantically or phonologically related context word affects word-production latencies, there are discrepant findings with respect to the polarity of these effects (facilitation or interference, relative to an unrelated control). Some tasks show semantic facilitation, but other tasks show semantic interference. Bloem and La Heij (2003), following Neumann (1986), called this the semantic relatedness paradox. For example, picture–word interference experiments (e.g., A. S. Meyer & Schriefers, 1991; Schriefers et al., 1990; Starreveld & La Heij, 1995), which require picture naming in the context of distracter words, typically show semantic interference, provided the distracter is presented briefly before or simultaneously with the target. However, other paradigms show semantic facilitation (e.g., Bloem & La Heij, 2003; Bloem, Van den Boogaard, & La Heij, 2004; Roelofs, 2003). This is the case when the picture is a distracter and the target is a word to be used in a phrase (Roelofs, 2003) or a word that is to be translated (Bloem & La Heij, 2003). It is also the case when the distracter is a word presented relatively long (400 ms) before the target (Bloem et al., 2004; also see Alario, Segui, & Ferrand, 2000, Footnote 1; Glaser & Düngelhoff, 1984).

Bloem et al. (2004) explained this semantic paradox as the result of interplay of activation at the levels of lexical concepts and lemmas (see Figure 1). A given target concept will be lexicalized faster if a semantically related concept is active, as a result of spreading activation at the conceptual level. However, the target lemma will be lexicalized more slowly if a semantically related lemma is active, as a result of competition at the lemma level (Roelofs, 1992; Starreveld & La Heij, 1996). If the distracter is a picture, its concept, but not its lemma, will be activated, leading to facilitation. Bloem et al. proposed that there is slower decay at the conceptual level than at the lemma level. Therefore, a distracter word presented relatively long ago will be a weak competitor at the lemma level but will still prime the target at the conceptual level,[1] whereas a distracter word presented just before the target will be a strong competitor at the lemma level. Thus, the polarity of semantic effects depends on the time interval between distracter and target.[2]

Now consider the case in which speakers produce a word in error and then repair it with a semantically related word (e.g., "strawberry . . . grape"), for example in picture naming. This is a situation in which there may be residual activation for both the lexical concept and the lexical representation of *strawberry* at the moment the speaker intends to retrieve the word *grape*. One may, therefore, compare it with the situation just described above, in which a distracter word is presented prior to presenting a target (both the distracter's lexical concept and its lexical representation are active). As discussed above, the polarity of semantic effects in such situations is likely to vary with the time interval between distracter (error) and target (repair). Of course, in speech error repair, this time interval is not determined by external stimuli but rather by the subsequent intentions to say a word and to repair the mistaken production of that word. A rough estimate of that interval is the time between error onset and repair onset. This interval was 618 ms on average in a corpus of experimentally elicited repairs (Oomen & Postma, 2001) analyzed in Hartsuiker and Kolk (2001).[3] This is a long interval in comparison with the literature on context effects in word production; if the findings of Bloem et al. (2004) generalize to the situation of repair, one might predict that a semantic relation between error and repair will facilitate repair production. However, this interval is highly variable, and a considerable proportion of words is interrupted word-internally. If our claim is right that there is closer temporal synchrony between retrieval of the error and repair in such interrupted cases, there may be less semantic facilitation (or even inhibition) on the repair

[1] Bloem et al. (2004) made several further assumptions to account for the paradox, for instance, that only an intended concept will activate the lemma level and that intended concepts will activate a cohort of semantically related lemmas. A discussion of this theory is beyond our immediate purposes.

[2] As noted by Alario et al. (2000, p. 747), for long stimulus onset asynchronies (i.e., 400 ms or more), facilitation seems to be the rule. However, there is an exception to this generalization: Some studies have shown semantic interference when extremely long intervals (4 s and more) are used between a prime (the definition of a word) and a picture target (Vitkovitch, Rutter, & Read, 2001; Wheeldon & Monsell, 1994). Given this very different time course from "online" context effects, like picture–word interference, it is conceivable that in these studies a different set of processes comes into play.

[3] The mean error-to-repair interval is based on a corpus of 98 experimentally elicited repairs gathered in an experiment in which speakers described pictures in a network, resulting in utterances such as "from the red mailbox go straight up to the yellow uh green book." All were incidents with a single error and repair, and all were repairs of errors.

production for word interruptions than for uninterrupted error productions.

Overview of the Experiments

To test the effects of residual activation of an abandoned word's lexical representation on the planning of the subsequent word, we report two picture-naming experiments below. We used a paradigm inspired by Van Wijk and Kempen (1987), in which speakers begin naming one picture (the context picture), but that picture is soon after replaced by another picture (the target picture) on a small proportion of trials. When this happened, speakers were supposed to abandon their initial response to the context picture and name the target picture as fast and accurately as possible. We expected this to lead not only to responses in which participants managed to completely say the name of the context picture ("strawberry . . . grape") but also to responses in which the context name was interrupted ("straw— . . . grape"). Note that the majority of trials were no-change trials, so that participants would not anticipate the change. We also included unrelated change trials (fillers), so that participants would be less likely to discover a relationship between the pictures in the related change trials. In the change trials, the stimulus onset asynchrony was −300 ms (i.e., context picture 300 ms before target picture). On the basis of pilot experiments, this value was chosen so that on many occasions, the participants could not stop themselves from saying at least part of the context picture.

As mentioned above, our experimental situation is in some respects different from that of naturalistic speech error repair and is more similar to other naturalistic situations in which words are replaced. For example a sportscaster might begin to name a particular soccer player who has the ball and then realize that the identification had been incorrect. However, it is our contention that the experimental task has important commonalities with error repair, because it presents the speaker with the same situation as the one of having just detected an error: the need to abandon one word and say another word instead. At the same time, our task makes it possible to exert tight experimental control over the properties of the abandoned word (the error) and the one that replaces it (the repair) and the time course during which the replacement word becomes available.

Because we are interested in the processes following error detection, it is irrelevant for our purposes that in natural speech error repair, speakers can detect an error by inspecting their internal or external speech (Levelt, 1983), whereas in our experiments, an error is detected through visual perception of a new picture, which shows a different object from the picture currently being named. For the same reason, it does not matter that in normal speech production, speakers might first evaluate whether the error is so disruptive to communication that it is necessary to interrupt (Berg, 1986, 1992). Nevertheless, it is an interesting question to ask how the participants detect the need to abandon one word and produce another word instead in this paradigm. In one view, they exploit the minimally necessary information, namely, a change in low-level visual information when the target picture replaces the context picture. However, in another view, the speaker waits until the object is recognized (so that the corresponding lexical concept is activated) and has ensured that this corresponds to a different concept from the one currently being lexicalized. In the former view, amount of visual overlap between the pictures should be a determinant of whether the context name is interrupted or said completely, whereas a semantic relation between the names would be irrelevant. In the latter view, amount of visual overlap should be unimportant, but speakers would interrupt less often with semantically related pairs. This is because for these pairs, it would take longer to establish that the object to be named (the target) is different from the one currently being named (the context) because the two concepts are similar.

Experiment 1 evaluated the effects of semantically related target pictures. As mentioned above, according to a wipe-clean hypothesis, in which there is a sustained inhibition signal to the representations of the abandoned word, relatedness should not matter. But if representations of the abandoned word remain active, we predict semantic facilitation for completed productions (because of the relatively long time course between the two productions) but reduced semantic facilitation, or even semantic interference, for interrupted productions. We derive and test predictions for phonological relatedness in Experiment 2.

Experiment 1: Semantic Repairs

Method

Participants. Thirty-two students at the University of Edinburgh participated in this experiment. All were native speakers of English.

Materials. Critical items were 48 line drawings of common objects from the Snodgrass and Vanderwart (1980) set and similar sets, such as the picture database collected at the Max Planck Institute in Nijmegen (see Severens, Van Lommel, Ratinckx, & Hartsuiker, 2005). Twenty-four pictures had polysyllabic names (serving as context pictures) and 24 had monosyllabic names (serving as target pictures). By choosing context pictures with relatively long names, and target pictures with relatively short names, we intended to maximize the opportunity for within-word interruptions (cf. Brédart, 1991). Pictures were assigned to pairs, so that 24 semantically related pairs (i.e., pairs from the same categories) were created. A reassignment of the pictures led to 24 unrelated pairs. For any given context picture, the target pictures were matched for frequency according to the CELEX database (Baayen, Piepenbrock, & Van Rijn, 1993). Appendix A lists all critical picture names, and the first column of Table 1 lists the mean frequency and length (in number of phonemes and syllables) of the context and target pictures.

The task (see below) involved two successive picture presentations. Because it is conceivable that the amount of visual overlap between the pictures might influence some of the results, we determined the Euclidean Overlap (EO) metric (Laws & Gale, 2002) for the picture pairs. This metric is calculated by scaling the pictures to a square matrix of $n \times n$ pixels and by comparing the value (black is 1 and white is 0) of each pixel in a given picture with that in a second picture and then taking the square root of the summed squared difference. Given that the pictures subtend a matrix of $n \times n$ pixels, the EO is between 0 (perfect overlap) and n (no overlap). Laws and Gale (2002) determined this measure for almost every picture in the Snodgrass and Vanderwart (1980) set with every other picture, and they first scaled the pictures so that the boundaries of each picture would touch the boundary of the $n \times n$ pixel grid before calculating EO. This scaling method means, for example, that a large bike tube and a small bike tube of the same shape would overlap very strongly (small EO). These measures were available for most of our picture pairs (46 of 48 comparisons), and they are reported in Table 1 as the "scaled visual overlap." However, because a more critical variable for our purposes may be the number of pixels changing from context to target, we reimplemented Laws and Gale's algorithm but without first scaling the pictures. Thus, according to this unscaled metric there would be very little visual overlap between the

Table 1
Characteristics of the Stimuli in Experiments 1 and 2

Stimulus characteristics	Experiment 1	Experiment 2
Context		
Frequency (per 17.9 million)	629	636
Length (syllables)	2.33	2.21
Length (phonemes)	6.13	5.54
Target		
Frequency (per 17.9 million)**	9,574	1,516
Length (syllables)*	1.00	1.21
Length (phonemes)	3.17	3.46
Scaled visual overlap		
Related	11.5	11.4
Unrelated	11.7	11.1
Unscaled visual overlap		
Related	70.5	73.9
Unrelated	72.6	74.1

Note. Because the related and unrelated conditions were created by re-pairing the stimuli, the values for frequency, length in phonemes, and length in syllables are always identical between conditions.
* $p < .01$. ** $p < .001$. (Indicates significant differences across experiments by use of independent t tests.)

pictures of a small and large bike tube. We calculated this measure for all our picture pairs, and the data are reported in Table 1 as the "unscaled visual overlap." There was a large and significant correlation between these two measures ($r = .77$, $N = 46$, $p < .01$). Independent-sample t tests revealed that the visual similarity for related pairs did not differ reliably from that of the unrelated pairs: scaled visual overlap, $t(44) = 0.40$, $p = .69$; unscaled visual overlap, $t(46) = 0.53$, $p = .60$.

We assigned a further 24 pictures to 12 unrelated context-target pairs to reduce the proportion of related pictures in the change trials. Finally, 39 pictures served as the targets in the no-change trials.

List construction. Two basic lists were created, each consisting of 228 no-change trials (86.4% of total trials), obtained by presenting the 39 filler pictures four times and the 72 pictures for the change trials once. Additionally, there were 36 change trials (13.6% of total trials), 12 of which were fillers and 24 of which were critical trials. The context and target pictures were semantically related in 12 critical trials and unrelated in the other 12 critical trials. Across the two lists, each context picture appeared once with a related target picture and once with an unrelated target picture. The two lists had the same pseudorandom order, with the constraints that change trials were at least 5 and at most 9 trials apart and that the pictures in a change trial were at least 20 trials apart from the same pictures in a no-change trial (with half the pictures appearing first in the change trial and half in the no-change trial). To further control for any order effects, we derived two further lists by reversing the order of the basic lists. Each of the four lists was presented to 8 participants.

Procedure. The participants were first familiarized with the names of all 111 pictures. Each picture was presented with its name printed underneath. Participants were asked to name the picture aloud and progressed to the next picture by pressing a key. Subsequently, participants were again asked to name all pictures, this time without the name being provided.

In the experimental phase, participants were instructed to name all the pictures that appeared on the screen as quickly as possible. If one picture appeared immediately after another, participants were to try to stop naming the first picture and name the second picture as quickly as possible. In all trials, a fixation cross appeared in the center of the screen for 1,000 ms, followed by a blank screen for 500 ms. In the no-change trials, the target picture subsequently appeared for 500 ms. In the change trials, the context picture appeared for 300 ms and was replaced by the target picture, which appeared for 500 ms.

Apparatus. The experiment was implemented in E-prime 1.0 (Scheider, Eschman, & Zuccolotto, 2002), and speech was recorded on digital audiotapes (DATs) with a high-quality directional microphone. The DATs were transferred to a computer file. Latencies between appearance of the target picture and onset of its name were manually determined with the phonetic software package PRAAT (Boersma & Weenink, 1992). We also recorded latencies between onset of context picture and its name as well as naming latencies for the same pictures in no-change trials.

Scoring and data analysis. Because of equipment failures, 10 responses (1.3%) were discarded from analysis. A further 60 responses were discarded because of errors on the target picture. There was no effect of relatedness on the number of errors in naming the targets (unrelated correct = 349, error = 32; related correct = 349; error = 28; $\chi^2 < 1$).

Finally, 42 responses were discarded because either the context picture was described with a different word than we intended or the context or target picture was named in error when it occurred as a no-change filler. The resulting 656 responses were divided into three categories: interrupted context word (e.g., "straw— . . . grape"), complete context word (e.g., "strawberry. . . grape"), and context word skipped (e.g., ". . . grape").

The analyses treated relatedness and response type as within-participant and within-item variables. Because response type was not under experimental control, and there turned out to be highly divergent numbers of observations per cell, traditional analyses, one averaging over items and one over participants, were not appropriate. We therefore opted to analyze this design using linear mixed-effect (multilevel) models (Pinheiro & Bates, 2000), so that every observation counted, and one can simultaneously generalize to participants and items.

Results

The number of responses of each type is listed in Table 2. Whereas interrupted context words were more frequent in the unrelated condition, completed context words occurred more often in the related condition. Skipped context words occurred roughly equally often. The difference in response frequency by condition was significant, $\chi^2(2, N = 656) = 6.48$, $p < .05$. To explore this difference, we submitted the proportions of interrupted, completed, and skipped context picture names for the related and unrelated conditions to paired-sample t tests. There were significantly more completed responses in the related condition than in the unrelated condition, $t(31) = 2.90$, $p < .01$. The relatedness effect was not significant for the interrupted responses, $t(31) = 1.80$, $p = .08$, or the skipped responses, $t(31) = 1.61$, $p = .12$. Finally, the correlations between visual similarity and the proportion of completed responses approached significance (scaled visual overlap: $r = -.27$, $N = 46$, $p = .08$; unscaled visual overlap: $r = -.27$, $N = 48$, $p = .06$).

The results indicated that speakers found it more difficult to interrupt a context word if it was semantically related to the target. This finding cannot be attributed to the participants finding it more

Table 2
Frequency of Each Response Type for the Unrelated and Semantically Related Conditions in Experiment 1

Response	Unrelated	Related
Context interrupted	110	87
Context completed	114	142
Context skipped	108	95

Note. Responses containing errors were excluded.

Table 3
Mean Naming Latency (in Milliseconds) for the Context and Target Pictures in Experiment 1 by Response Type and Relatedness

	Unrelated		Related				
Response	*M*	*SD*	*M*	*SD*	*M*	*SD*	Difference
Context naming latency							
Context interrupted	591	151	611	162	600	156	−20
Context completed	599	158	585	136	591	146	14
M	595	155	595	147			
Target naming latency							
Context interrupted	676	167	729	202	700	185	−53
Context completed	761	201	716	188	736	195	45
Context skipped	744	167	765	205	754	186	−21
M	727	183	734	197			

Note. *Difference* refers to the difference between the unrelated and related conditions. Target naming latencies were measured relative to the onset of the target picture.

difficult to detect a low-level visual change in the semantically related pairs: The marginally significant correlations between visual overlap and the probability of completing the context word suggested there could be some role of visual factors, but the visual overlap did not differ between the related and the unrelated pairs. Thus, the difference in likelihood of completing cannot be attributed to that variable. Rather, the data suggest an effect of semantic relatedness on the probability of completing over and beyond any effect of visual overlap, consistent with the suggestion in the introduction that the speaker postpones interruption until it is established that the target object is not the same as the context object.

We considered response times more than three standard deviations from the grand mean as outliers and excluded them from analysis (1.7%). The mean naming latencies in the remaining data are listed in Table 3.

An initial analysis on the response latencies for the target pictures revealed that there were no main effects of relatedness or response type (Fs < 1) but that the Relatedness × Response Type interaction was significant, $F(2, 608) = 4.28$, $p < .02$. Separate analyses for each response type showed that the 53-ms semantic interference effect for the interrupted context word responses was significant, $F(1, 163) = 5.35$, $p < .05$, and that the 45-ms semantic facilitation effect in the completed context word responses approached significance, $F(1, 218) = 3.61$, $p = .06$. The 21-ms semantic interference effect in the context word skipped responses was not significant ($F < 1$).

Because not every item yielded the same number of interrupted-context and completed-context responses, it was important for us to statistically control for any effects of item differences. We therefore conducted additional analyses that treated log frequency of both the context and the target picture names, unscaled visual overlap between context and target picture, and the length in phonemes of the context picture name as covariates.[4] Both frequency and length may affect response latencies (Jescheniak & Levelt, 1994; A. S. Meyer, Roelofs, & Levelt, 2003), and an analysis of the no-change trials in which the target pictures were named in isolation showed that frequency predicted the onset times, $F(1, 612) = 5.55$, $p < .02$. As before, there were no main effects of relatedness or response type (Fs < 1), but the interaction was significant, $F(2, 604) = 5.08$, $p < .01$. A separate analysis for each response type revealed that for interrupted-context responses, the semantic interference effect was significant, $F(1, 159) = 6.39$, $p < .02$, and that in the completed-context responses the semantic facilitation effect was significant, $F(1, 214) = 5.32$, $p = .02$, but that for the skipped-context responses there was no semantic effect ($F < 1$).

An analysis on onset latencies for the context pictures (obviously restricted to incidents in which the context name was produced partially or completely) showed, importantly, that there was no effect of relatedness ($F < 1$) and no interaction between relatedness and response type, $F(1, 406) = 1.99$, $p = .16$. Thus, the response time for the context picture did not depend on the relation between context and target word. However, there was a main effect of response type that approached significance, $F(1, 406) = 3.63$, $p = .06$, suggesting that speakers were faster to initiate articulation of the context name when they subsequently completed it (591 ms) than when they subsequently interrupted it (600 ms).

Discussion

As expected, participants often could not prevent themselves from saying the context word (on two thirds of the critical trials), although they did manage to interrupt within the context word on a substantial number of trials. The time course of repair was affected by semantic relatedness. There was a full reversal of the polarity of these effects, depending on whether speakers interrupted the context name: There was semantic interference for the interrupted-context productions, and there was semantic facilitation for the complete-context productions. Additionally, participants apparently found it more difficult to interrupt in the semantically related condition than in the unrelated condition.

Experiment 1 showed evidence for the semantic paradox in a single experiment. However, note that there is also a phonological

[4] An analysis with scaled visual overlap yielded the same pattern of results. Length of the target picture did not vary enough to be meaningfully treated as a covariate.

relatedness paradox in the literature: Some tasks show phonological facilitation, but other tasks show phonological interference (Sevald & Dell, 1994; Wheeldon, 1999, 2003). Experiment 2 tested whether the production of a target picture's name was affected by participants just having abandoned a phonologically related name.

Experiment 2: Phonological Repairs

This experiment tested the effects of phonological relatedness (i.e., overlap in the initial phoneme or phonemes) between error and repair on repair production. As in Experiment 1, we derive predictions for this situation from the literature on context effects in word production. These studies provide evidence for a phonological paradox: Naming a picture of a fork is facilitated by hearing the distracter word *folk* (e.g., Schriefers et al., 1990), but naming the same picture is inhibited after having said "folk" in response to a definition (e.g., Wheeldon, 2003). As pointed out by O'Seaghdha, Dell, Peterson, and Juliano (1992), there is also a complex pattern of facilitation and interference in studies on word naming that depends on item and task variables (e.g., prime duration, prime frequency, and target frequency; see also Lupker & Colombo, 1994; O'Seaghdha & Marin, 2000).

In word production studies, phonological facilitation is usually found in tasks in which the context word should be ignored, such as picture–word interference (e.g., A. S. Meyer & Schriefers, 1991; Schriefers et al., 1990; Starreveld, 2000). Additionally, phonological relatedness helps if speakers produce words in blocks that are homogeneous with respect to the initial phoneme(s), at least as long as the words all share the same word frame (i.e., the number of syllables and the stress pattern). This is most obvious in the implicit priming paradigm (A. S. Meyer, 1990, 1991; Roelofs & Meyer, 1998), in which participants first learn to associate word pairs and then name the second word of each pair upon presentation of the first word as a cue. However, effects of block homogeneity on phonological relatedness effects also have been observed in picture naming (Damian, 2003; Roelofs, 1999).

In contrast, there is phonological interference when both the context and target words are produced overtly and the blocks are not homogeneous (O'Seaghdha & Marin, 2000; Rogers & Storkel, 1998; Sevald & Dell, 1994; Sullivan & Riffel, 1999; Wheeldon, 2003) and when speakers plan to say the context word aloud but are occasionally required to produce another target (e.g., D. E. Meyer & Gordon, 1985; O'Seaghdha et al., 1992; Yaniv, Meyer, Gordon, Huff, & Sevald, 1990). It is important to note, however, that although there are consistent phonological interference effects when the words have begin overlap (i.e., have the same initial phonemes), this is not so when the words have end overlap (i.e., have the same final phonemes; see O'Seaghdha & Marin, 2000; Sevald & Dell, 1994). Some authors even reported facilitation in the latter case (Sevald & Dell, 1994).

Levelt et al. (1999) attributed the phonological facilitation effect in picture–word interference studies to priming at the level of phoneme representations. Because of this priming, speakers are faster to retrieve the correct phoneme. The facilitation in implicit priming and related tasks would have a different locus. According to Levelt et al., the homogeneity benefit results from preparation at the level of the phonological word. Speakers in a homogeneous block would prepare as much as they can of the word frame with as many phonemes as possible assigned to initial position(s). The reasons for postulating these two different loci of phonological effects are discussed at length in Levelt et al.'s study.

However, Levelt et al.'s (1999) model, which was designed only to account for single-word utterances, does not explain the phonological interference effects observed when speakers say one word and then say a phonologically similar word (e.g., Sevald & Dell, 1994). In contrast, Sevald and Dell proposed the competitive-cuing model. In that model, phonemes are activated sequentially, and activation flows bidirectionally between words and phonemes. This model accounts for the interference when the words have begin overlap. Suppose a speaker has just said "pick" and now says "pin." After activating the first phoneme of *pin*, /p/, the lexical representation of *pick* becomes active again, and it becomes even more active after the second phoneme of *pin*, /i/, is activated. As a result, *pick*'s final phoneme, /k/, becomes active too and competes with the last phoneme of *pin*, /n/, thereby delaying selection. However, if a speaker has just said "pick" and now says "tick," there is no disadvantage. The /t/ of *tick* will not activate the lexical representation for *pick*, and although the /i/ of *tick* activates *pick*, this will only cue the correct final phoneme /k/. This predicts facilitation in the case of end overlap.

The situation in which a speaker produces a word in error and then repairs the error with a phonologically related word (e.g., "lemon . . . leg") is an example of a two-word utterance with phonological overlap. Given the many earlier studies showing inhibition in the case of begin overlap, and the competitive-cuing model's hypothesis that shared phonemes at the beginning of the target word will reactivate the context word, thus leading to a miscuing of the final phonemes, the most obvious prediction is that the error would interfere with the repair in this situation. The situation may be different, however, for situations in which the speaker interrupts word-internally. For example, in the case of the picture pair *lemon–leg*, the speaker might produce "le— . . . leg." In that situation, the mismatching phonemes have not been uttered, and it is conceivable that they will be relatively less activated, even though the context word is reactivated by the target. In that case, cuing of the matching phonemes may even lead to facilitation.

Method

Participants. Thirty-two students at the University of Edinburgh participated. All were native speakers of English, and none of them had participated in Experiment 1.

Materials. Critical items were 56 line drawings of common objects from the Snodgrass and Vanderwart (1980) and similar sets. Twenty-eight pictures had polysyllabic names (serving as context picture), and 28 had mostly monosyllabic names (target picture), but the constraints on material selection forced us to include 5 disyllabic pictures. Pictures were assigned to pairs, so that 28 phonologically related pairs were created (sharing the first or the first two phonemes). As mentioned above, manipulations of begin overlap have shown the clearest effect in the word-production literature.

A reassignment of pictures led to 28 unrelated pairs. For any given context picture, the pictures serving as target were matched for frequency according to the CELEX database. Appendix B lists all critical picture names, and the second column of Table 1 shows the length, frequency, and degree of visual overlap of these stimuli. Scaled visual overlap was available for 46 of the 56 comparisons. As in Experiment 1, there was a large and significant correlation between scaled and unscaled visual overlap ($r = .72$, $N = 46$, $p < .01$). There were no reliable differences between

Table 4
Frequency of Each Response Type for the Unrelated and Phonologically Related Conditions in Experiment 2

Response	Unrelated	Related
Context interrupted	63	67
Context completed	281	254
Context skipped	36	36

Note. Responses containing errors were excluded.

the amount of overlap in the two conditions: scaled visual overlap, $t(44) = 0.52$, $p = .60$; unscaled visual overlap, $t(54) = 0.04$, $p = .97$.

We assigned a further 28 pictures to 14 unrelated context–target pairs, to reduce the proportion of related pictures in the change trials. Additionally, 40 pictures served as the targets in the no-change trials.

List construction. Two basic lists were created, each consisting of 244 no-change trials (85.3% of total trials), which we obtained by presenting the 40 filler pictures four times, and the 84 pictures, also used in the change-trials, once. Additionally, there were 42 change trials (14.7% of total trials), 14 of which were fillers and 28 of which were critical trials. The context and target stimuli were related in 14 critical trials and unrelated in the other 14 trials. Across the two lists, each context picture appeared once with a related and once with an unrelated target. The two lists had the same pseudorandom order, with the constraints that change trials were at least three and at most eight items apart and that the pictures in each change trial were at least 20 trials apart from the same pictures in a no-change trial (with half the pictures appearing first in the change trial and half in the no-change trial). To further control for order effects, we derived two further lists by reversing the order of the basic lists. Each of the four lists was presented to 8 participants.

Procedure and apparatus. These were identical to Experiment 1.

Scoring and data analysis. The same scoring regime and data analyses were used as in Experiment 1. We discarded 160 responses (17.9%) because of errors in the name of the context picture (19), the target picture (44), or disfluencies in the context picture (97). The target picture was incorrectly named more often in the phonologically related condition (correct = 419, error = 29) than in the unrelated condition (correct = 433, error = 15), $\chi^2(1, N = 896) = 4.04$, $p < .05$. A further subdivision showed that most of these errors (33) occurred for context-completed responses, and of these 33, there were significantly more errors in the related condition (correct = 254, error = 22) than in the unrelated condition (correct = 281, error = 9), $\chi^2(1, N = 566) = 6.47$, $p < .05$. Twelve of the remaining responses (1.6%) were excluded because they had target naming latencies of more than three standard deviations from the mean.

Results

The number of responses of each type is listed in Table 4. The difference in response frequency by condition was not significant ($\chi^2 < 1$). A comparison of the proportions of context interrupted, context completed, and context skipped responses for each condition revealed no significant effects (ts < 1).

The response latencies for both context and target stimuli are listed in Table 5. An initial analysis on the response latencies for the targets revealed that there was a main effect of relatedness, $F(1, 688) = 4.50$, $p < .05$; no effect of response type, $F(2, 688) = 1.81$, $p = .17$; and a significant interaction, $F(2, 688) = 3.86$, $p < .05$. Separate analyses for each response type showed that the 39-ms phonological facilitation effect in the interrupted-context responses approached significance, $F(1, 109) = 3.18$, $p = .08$, and that the 465-ms phonological interference effect for the completed-context responses was significant, $F(1, 493) = 9.98$, $p < .01$. There was no relatedness effect for skipped-context responses ($F < 1$).

As in Experiment 1, additional analyses included log frequency of the context and target pictures, unscaled visual overlap, and length of context and target picture names in phonemes as covariates. The pattern of results remained the same: relatedness, $F(1, 683) = 4.31$, $p < .05$; response type, $F(2, 683) = 1.45$, $p = .23$; Relatedness × Response Type, $F(2, 683) = 3.92$, $p = .02$. A separate analysis for each response type revealed that for completed-context responses, the phonological interference effect was significant, $F(1, 488) = 9.24$, $p < .01$. For the interrupted-context responses, the phonological facilitation effect remained marginally significant, $F(1, 104) = 3.00$, $p = .09$, and for the skipped-context responses, there was no relatedness effect ($F < 1$).

An analysis on context picture-naming latencies including frequency and length of context and target picture names as covariates demonstrated an effect of response type, $F(1, 615) = 6.38$, $p < .02$. Speakers were 9 ms faster to initiate articulation of the context name when they subsequently interrupted it (544 ms) than when they subsequently completed it (553 ms). Although descriptively this effect appears to be primarily driven by the related

Table 5
Mean Naming Latency (in Milliseconds) for the Context and Target Pictures in Experiment 2 by Response Type and Relatedness

	Unrelated		Related				
Response	*M*	*SD*	*M*	*SD*	*M*	*SD*	Difference
Context naming latency							
Context interrupted	549	114	539	96	544	105	−10
Context completed	546	117	561	130	553	124	−15
M	546	117	557	124			
Target naming latency							
Context interrupted	660	162	621	155	640	159	39
Context completed	652	139	698	169	674	156	−46
Context skipped	753	169	731	142	742	155	22
M	663	148	687	167			

Note. *Difference* refers to the difference between the unrelated and related conditions. Target naming latencies were measured relative to the onset of the target picture.

condition (22 ms) rather than the unrelated condition (−3 ms), there was no relatedness effect, $F(1, 615) = 1.35$, $p = .24$, and no interaction ($F < 1$). The effect of relatedness was not significant for either interrupted-context ($F < 1$) or completed-context responses, $F(1, 489) = 1.53$, $p = .22$. Note that there was a tendency in the opposite direction (also of 9 ms) in Experiment 1.

Discussion

Again, participants could usually not stop themselves saying at least part of the context word. In contrast to Experiment 1, however, the majority of responses now included cases in which the context word was completely produced. We do not know what caused this difference, but because this also held for unrelated targets, whereas there were many more interrupted context responses with unrelated targets in Experiment 1, the difference between the two experiments is probably due to differences between the two groups of participants and the two sets of context items and not to an effect of experimental manipulations.

There was no difference between the phonologically related and unrelated conditions in terms of the distribution of responses, which fits the account that error detection is affected only by conceptual similarity. Phonologically related pairs are as different conceptually as unrelated pairs.

As predicted, the data showed an overall phonological interference effect: Participants named the target more slowly if it was phonologically related to the context word. However, as in Experiment 1, there was an interaction with response type, and separate analyses suggested that, again, there was a reversal of polarity: There was phonological interference in the completed-context responses, but there was a strong tendency toward phonological facilitation in the interrupted-context responses.

General Discussion

In summary, two experiments tested whether the representations of abandoned words remain active and can influence the production of words that are planned to replace them. Speakers were presented with a picture to name, and while they began to name it, the picture was replaced with a semantically or phonologically related picture. The data showed that speakers usually could not stop themselves saying the name of the context picture, at least partly; that speakers were more likely to completely say that name if the two pictures showed objects from the same conceptual category; and, crucially, that the time to say the target picture was a function of the type of relation with the prime picture (semantic or phonological) as well as a function of whether the name of the prime picture was said completely or was interrupted.

Our experiments suggest that the production of repairs is affected by the relation between error and repair: There remains residual activation of the error's (sub)lexical representations, and this activation influences production of the repair. This held both for the situation in which participants interrupted the production of the error and for cases in which they said the error completely (although the polarity of effects was different for these two situations; see below). Thus, our results argue against the hypothesis that an interruption signal sent to every language production component results in a wiping clean of the activation of all relevant lexical and sublexical representations. Of course, this does not necessarily mean that speakers interrupt without sending an interruption signal to every component. It implies that when an interruption signal is sent, it does not reset the activation of the error's representations to resting level. It is possible, for example, that the error's representations are briefly inhibited but that the activation soon rises again (see Dell, 1986, for a computational model in which such a rebound follows a brief inhibition subsequent to normal selection processes). However, it is also possible that the stop signal would affect whether representations are selected for production or not. For example, in Levelt et al.'s (1999) model (also see Roelofs, 1997), a connectionist architecture of the lexicon is supplemented by a rule-based system. These rules ensure that only representations corresponding to the intended word will be selected. Thus, on such an architecture, the activation in the lexical network remains intact, whereas the abandoned word is no longer selected for production.

The experiments also showed that speakers often found it difficult to prevent themselves from saying the error at least partly; in many cases, they even did not manage to interrupt word-internally. Nevertheless, presentation of the target pictures occurred already 300 ms after that of the context pictures, which left, on average, 300 ms before the context picture was named overtly. Given the estimate that it takes about 200 ms to stop speech (Ladefoged, Silverstein, & Papçun, 1973), our speakers did not appear to begin interrupting themselves as soon as they detected a change in the visual display (i.e., exploiting the minimally necessary information). That suggestion is strongly supported by the finding that the probability of interrupting was a function of the relation between context and target picture: It was more difficult to interrupt if the two pictures were semantically related than if they were unrelated (Experiment 1), but it was not more difficult to interrupt if the two pictures were phonologically related than if they were unrelated (Experiment 2).

We suggest, following Hartsuiker and Kolk (2001), that the processes of repair and interruption are triggered only after a process of comparison has taken place at the conceptual level (Levelt, 1989). In the case of a semantic error, this process takes longer, because the two concepts are more similar than in the case of an unrelated error. As a result, the process of interruption starts somewhat later, and it is more likely that the error is said completely. Our claim is, therefore, that the moment of error detection, relative to the moment when the end of the word is articulated, is the deciding factor in determining whether a word is interrupted, in agreement with Brédart (1991) and Levelt (1989).[5]

Perhaps most important, our experiments showed that relatedness between error and repair could both help and hinder the production of the repair. In the situation that is most common in naturalistic speech error repairs, namely when the error is completely said (Brédart, 1991; Levelt, 1983), a semantic relation helped the production of the repair, but a phonological relation hindered. However, when the speaker managed to interrupt naming the error, there was a complete reversal of the polarity of both effects: semantic interference and phonological facilitation. To our

[5] Of course, another determinant is whether the utterance was an error or whether it was merely inappropriate. It is important to note, though, that Brédart (1991) demonstrated that inappropriate words are also more likely to be interrupted if they are longer.

knowledge, this study is the first to observe both the semantic and phonological relatedness paradoxes in a single paradigm.

In the introduction, we derived predictions about semantic context effects in speech error repair on the basis of current models of word production. We predicted that when the context word is said completely, a semantic relation would facilitate the production of the target because of the relatively long delay between retrieval of the context item's name and the target item's name. However, in the case of interrupted-context responses, with closer temporal synchrony, we predicted a reduction of facilitation or possibly even semantic interference. The data were in accordance with these predictions. We therefore follow the account provided by Bloem et al. (2004) that (a) pictures do not automatically activate their names and that (b) there is faster decay at the lexical than at the conceptual levels. With respect to our task, the first assumption means that the mere presentation of the target picture does not imply that its name retrieval already begins, and the second assumption means that the polarity of semantic effects is a function of time course. As a result of the close temporal synchrony between retrieving context and target in the case of interrupted-context responses, residual activation at the lexical level hinders more than residual activation at the conceptual lexical helps. However, this reverses in the case of completed-context responses, in which lexical activation of the context word has had time to decay.

We also predicted that a phonological relation between context and target would interfere with the production of the target on the basis of Sevald and Dell's (1994) competitive-cuing model. In that model, a begin-overlap phonological relation between two words hinders the production of the second word, because the shared phonemes (at the beginning of the second word) reactivate the first word, thus activating the later phonemes, which are different between the words. The activation of these mismatching phonemes makes it more difficult to select the final phonemes of the second word. Thus, we predicted a phonological interference effect, and this prediction was confirmed for incidents in which the context name was completely said, which constituted the majority of cases in Experiment 2 and in corpora of naturalistic speech error repairs (e.g., Brédart, 1991).

There was also an interaction between response type and phonological relatedness: There was phonological interference for the completed context responses but (marginally significant) phonological facilitation for the interrupted-context responses. That latter finding is consistent with the competitive-cuing model, because in the case of interrupted-context responses ("le— . . . leg"), fewer (or none) of the mismatching phonemes have been uttered. It can thus be expected that even though the initial phonemes of *leg* reactivate the lexical representation of *lemon* (the interrupted word), this will lead to relatively weak activation of the mismatching phonemes /m/, /ə/, and /n/. In contrast, the matching phonemes /l/ and /e/ will be relatively active, thus tipping the balance toward facilitation rather than inhibition.

In conclusion, then, relatedness between the words speakers abandon and the words with which they replace them can sometimes help and sometimes hinder production. These data are compatible with recent views on word production, in particular the theories proposed by Bloem et al. (2004) on the interface between concepts and words and by Sevald and Dell (1994) on the interface between words and phonemes.

References

Alario, F. X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53A,* 741–764.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., et al. (1991). The HCRC Map Task Corpus. *Language and Speech, 34,* 351–366.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database.* Philadelphia: Linguistic Data Consortium.

Berg, T. (1986). The aftermath of error occurrence: Psycholinguistic evidence from cutoffs. *Language and Communication, 6,* 195–213.

Berg, T. (1992). Productive and perceptual constrains on speech-error correction. *Psychological Research, 54,* 114–126.

Blackmer, E. R., & Mitton, E. R. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition, 39,* 173–194.

Bloem, I., & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language, 48,* 468–488.

Bloem, I., Van den Boogaard, S., & La Heij, W. (2004). Semantic facilitation and semantic interference in language production: Further evidence for the conceptual selection model of lexical access. *Journal of Memory and Language, 51,* 307–323.

Boersma, P., & Weenink, D. (1992). *Praat.* http://www.fon.hum.uva.nl/praat

Brédart, S. (1991). Word interruption in self-repairing. *Journal of Psycholinguistic Research, 20,* 123–138.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14,* 177–208.

Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 318–344.

Damian, M. F. (2003). Articulatory duration in single-word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 416–431.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 345–361.

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

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27,* 124–142.

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.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47,* 27–52.

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

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture–word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 640–654.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General, 118,* 13–42.

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science, 8,* 191–219.

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology, 42,* 113–157.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word-frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 824–843.

Ladefoged, P., Silverstein, R., & Papçun, G. (1973). Interruptibility of speech. *Journal of the Acoustical Society of America, 54,* 1105–1108.

Laws, K. R., & Gale, T. M. (2002). Category-specific naming and the “visual” characteristics of line drawn stimuli. *Cortex, 38,* 7–21.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14,* 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

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.

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition, 50,* 239–269.

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition, 7,* 485–495.

Lupker, S. J., & Colombo, L. (1994). Inhibitory effects in form priming: Evaluating a phonological competition explanation. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 437–451.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and 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 and Language, 30,* 69–89.

Meyer, A. S., Roelofs, A., & Levelt, W. J. M. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language, 48,* 131–147.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 1146–1160.

Meyer, D. E., & Gordon, P. C. (1985). Speech production: Motor programming of phonetic features. *Journal of Memory and Language, 24,* 3–26.

Neumann, O. (1986). *Facilitative and inhibitory effects of “semantic relatedness”* (Report No. 111/1986). Germany: University of Bielefeld.

Oomen, C. C. E., & Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research, 30,* 163–184.

O’Seaghdha, P. G., Dell, G. S., Peterson, R. R., & Juliano, C. (1992). Models of form-related priming in comprehension and production. In R. G. Reilly & N. E. Sharkey (Eds.), *Connectionist approaches to natural language processing* (pp. 373–408). Hove, England: Erlbaum.

O’Seaghdha, P. G., & Marin, J. W. (2000). Phonological competition and cooperation in form-related priming: Sequential and nonsequential processes in word production. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 57–73.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed effects models in S and S-Plus.* New York: Springer-Verlag.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition, 77,* 97–131.

Postma, A., & Kolk, H. H. J. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research, 36,* 472–487.

Postma, A., & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech, 39,* 375–392.

Rayner, K., & Posnansky, C. (1978). Stages of processing in word identification. *Journal of Experimental Psychology: General, 107,* 64–80.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42,* 107–142.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64,* 249–284.

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

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review, 110,* 88–125.

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.

Rogers, M. A., & Storkel, H. L. (1998). Reprogramming phonologically similar utterances: The role of phonetic features in pre-motor encoding. *Journal of Speech, Language, and Hearing Research, 41,* 258–274.

Scheider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime reference guide*. Pittsburgh, PA: Psychology Software Tools.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time-course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29,* 86–102.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition, 53,* 91–127.

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica, 119,* 159–187.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215.

Starreveld, P. A. (2000). On the interpretation of context effects in word production. *Journal of Memory and Language, 42,* 497–525.

Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 686–698.

Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 896–918.

Sullivan, M. P., & Riffel, B. (1999). The nature of phonological encoding during spoken word retrieval. *Language and Cognitive Processes, 14,* 15–45.

Van Wijk, C., & Kempen, G. (1987). A dual system for producing self-repairs in spontaneous speech: Evidence from experimentally elicited corrections. *Cognitive Psychology, 19,* 403–440.

Vitkovitch, M., Rutter, C., & Read, A. (2001). Inhibitory effects during object name retrieval: The effect of interval between prime and target on picture naming responses. *British Journal of Psychology, 92,* 483–506.

Wheeldon, L. R. (1999). Competitive processes during word-form encoding. *Behavioral and Brain Sciences, 22,* 59–60.

Wheeldon, L. R. (2003). Inhibitory form priming of spoken word production. *Language and Cognitive Processes, 18,* 81–109.

Wheeldon, L. R., & Monsell, S. (1994). Inhibition of spoken word production by priming a semantic competitor. *Journal of Memory and Language, 33,* 332–356.

Wilshire, C. E. (1998). Serial order in phonological encoding: An exploration of the “word onset effect” using laboratory-induced errors. *Cognition, 68,* 143–166.

Wilshire, C. E. (1999). The “tongue twister” paradigm as a technique for studying phonological encoding. *Language and Speech, 42,* 57–82.

Yaniv, I., Meyer, D. E., Gordon, P. C., Huff, C. A., & Sevald, C. A. (1990). Vowel similarity, connectionist models, and syllable structure in motor programming of speech. *Journal of Memory and Language, 29,* 1–26.

(*Appendixes follow*)

Appendix A

Critical Stimuli Used in the Change Trials of Experiment 1

Context	Unrelated target	Related target
artichoke	cup	corn
banana	flute	pear
butterfly	pan	ant
cigarette	tree	pipe
elephant	grape	bear
finger	couch	nose
flower	pipe	tree
guitar	knife	harp
kettle	ant	pan
ladder	cake	stairs
lemon	bear	grape
lobster	saw	fish
motorcycle	blouse	bus
necklace	swan	ring
paintbrush	cat	pen
peacock	ring	swan
bottle	corn	cup
sandwich	stairs	cake
scissors	harp	knife
screwdriver	fish	saw
jumper	bus	blouse
table	nose	couch
monkey	pen	cat
trumpet	pear	flute

Appendix B

Critical Stimuli Used in the Change Trials of Experiment 2

Context	Unrelated target	Related target
apple	monkey	axe
banana	scissors	balloon
barrel	tie	bag
bottle	hat	box
butterfly	pipe	button
candle	pig	cat
carrot	flag	camel
castle	eye	car
cigarette	balloon	scissors
finger	tree	fish
flower	camel	flag
glasses	whistle	glove
hammer	box	hat
iron	car	eye
ladder	pen	lamp
lemon	train	leg
mountain	spoon	mouse
mushroom	axe	monkey
penguin	lamp	pen
pillar	cat	pig
pineapple	button	pipe
radio	scarf	rake
skeleton	rake	scarf
spider	mouse	spoon
tiger	bag	tie
trousers	leg	train
trumpet	fish	tree
windmill	glove	whistle

Received September 16, 2004
Revision received April 13, 2005
Accepted April 15, 2005 ■