Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Memory and Language 52 (2005) 58–70

Journal of Memory and Language

www.elsevier.com/locate/jml

The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al. (1975)☆

Robert J. Hartsuiker[a,*], Martin Corley[b], Heike Martensen[c]

[a] Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium
[b] University of Edinburgh, UK
[c] University of Antwerp, Belgium

Received 3 June 2004; revision received 26 July 2004
Available online 1 October 2004

Abstract

The lexical bias effect is the tendency for phonological speech errors to result in words more often than in nonwords. This effect has been accounted for by postulating feedback from sublexical to lexical representations, but also by assuming that the self-monitor covertly repairs more nonword errors than word errors. The only evidence that appears to exclusively support a monitoring account is Baars, Motley, and MacKay's (1975) demonstration that the lexical bias is modulated by context: There was lexical bias in a mixed context of words and nonwords, but not in a pure nonword context. However, there are methodological problems with that experiment and theoretical problems with its interpretation. Additionally, a recent study failed to replicate contextual modulation (Humphreys, 2002). We therefore conducted two production experiments that solved the methodological problems. Both experiments showed there is indeed contextual modulation of the lexical bias effect. A control perception experiment excluded the possibility that the comprehension component of the task contributed to the results. In contrast to Baars et al., the production experiments suggested that lexical errors are suppressed in a nonword context. This supports a new account by which there is both feedback and self-monitoring, but in which the self-monitor sets its criteria adaptively as a function of context.
© 2004 Elsevier Inc. All rights reserved.

Keywords: Language production; Word production; Speech errors; Lexical bias effect; Feedback; Self-monitoring

Introduction

A debate that has occupied the language production literature for decades is whether the production of words should be considered a sequence of discrete levels (e.g., Levelt, Roelofs, & Meyer, 1999) or whether there is feedback between successive levels (e.g., Dell, 1986; see Rapp & Goldrick, 2000; Vigliocco & Hartsuiker, 2002 for reviews). An empirical finding that has often been used to argue for feedback is the *lexical bias effect*: the

☆ Parts of this work were presented at the 2002 AMLaP conference and the 2003 meeting of the Dutch Psychonomics Society. We thank Marieke Scholvinck, Nienke Korsten, and Hannah Furness for running the experiments. We also thank Jeff Bowers, Matthew Goldrick, Brenda Rapp, and an anonymous reviewer for their helpful comments on an earlier version of this manuscript. Rob Hartsuiker was supported by a grant from the Edinburgh University Development Trust Research Fund (E08679).
* Corresponding author. Fax: +32 9 264 6496.
E-mail address: Robert.Hartsuiker@ugent.be (R.J. Hartsuiker).

0749-596X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2004.07.006

tendency for phonological speech errors to result in existing words (e.g., *darn bore ⇒barn door*)[1] rather than nonwords (*dart board ⇒bart doard*), at a rate higher than predicted when phonemes were randomly exchanged between words (Baars, Motley, & Mackay, 1975; Dell, 1986, 1990; Dell & Reich, 1981; Hamm, Junglas, & Bredenkamp, 2004; Humphreys, 2002; Nooteboom, 2003, in press; but see Del Viso, Igoa, & García-Albea, 1991; Garrett, 1976). On the feedback account, phoneme representations prime representations for words, but not representations for nonwords, because the latter do not occur in the mental lexicon (by definition). If, for example, an unintended phoneme is highly active (e.g., the /d/ instead of /b/ when the target word is *bore)* feedback from the phonemes /d/, /o/, and /r/ will converge on the real word *door*, increasing the probability that it is accidentally produced.

However, the lexical bias effect is also compatible with discrete theories of language production, under the assumption that a self-monitoring system covertly filters out nonword errors more often than word errors (e.g., Levelt, 1989; Levelt et al., 1999; Nooteboom, in press; Roelofs, 2004). In particular, theories of verbal self-monitoring propose that the monitor inspects both external speech and a pre-articulatory representation, or "inner speech" (Hartsuiker & Kolk, 2001; Levelt, 1983). In order to account for the lexical bias effect they need to make two further assumptions. First, the self-monitoring system has to be fast enough to detect, and covertly and fluently repair, a phonological speech error before articulation. This assumption is supported (Motley, Camden, & Baars, 1982). Second, the monitor must use lexical status (*is this a word?*) as one of its criteria for monitoring, so that nonword errors are more readily detected than word errors. This second assumption has not been tested to date (but interestingly, Nooteboom, in press, recently showed that the monitor does not use lexical status in the correction of *overt* speech errors).

It has proven difficult to empirically distinguish between the feedback and monitoring accounts, since many findings are compatible with either account (see Hartsuiker, Schriefers, Bock, & Kikstra, 2003; Vigliocco & Hartsuiker, 2002 for discussion). For example, Dell (1986) showed a reduced lexical bias effect when speakers were required to speak more quickly. This is consistent with the feedback account (there is less time for feedback, thus less priming of lexical units), but it is equally consistent with a monitoring account (there is less time for covert repair, and thus less interception of nonword errors). However, one finding appears to support only the monitoring account, namely that lexical bias is modulated by context (Baars et al., 1975; Experiment 2). Baars et al. observed no lexical bias effect when the context consisted exclusively of nonwords, but there was a lexical bias effect when the context consisted of a mixture of words and nonwords. Since there is no obvious reason why feedback would vary with context, but there are clear reasons why a monitoring system would be flexible, adaptive, and hence sensitive to context, this finding has often been taken as the strongest evidence for the monitoring account (Levelt, 1989; Levelt et al., 1999; Nooteboom, in press; Roelofs, 2004).

Unfortunately, there are several methodological problems with Baars et al.'s (1975) experiment, and there is an important theoretical problem with the standard monitoring account (see below). Furthermore, Humphreys (2002) recently failed to replicate that experiment (albeit with a different procedure and different materials): She observed a lexical bias in both contexts. This article therefore asks whether the lexical bias is really mediated by context and whether it is mediated in the same way as the monitoring account predicts, namely by suppressing nonword errors in the mixed context.

Baars et al.'s experiments used the SLIP task, which involves the silent reading of briefly presented word or nonword pairs, occasionally followed by a cue to say the last pair aloud (a target pair). Typically, a target pair is preceded by several 'biasing items' that have a particular phonological relation with the target pair: The first phoneme of the first word in each biasing item is identical to the first phoneme of the second word in the target pair, and vice versa (see Table 1). The preceding context therefore primes the speaker to exchange the phonemes of the two target words.

In Baars et al.'s first experiment, all items were real words and when the initial consonants of the target words were exchanged they either resulted in real words or nonwords (Table 1). The results were clear-cut: Exchanges yielding a word outcome occurred more frequently than exchanges yielding a nonword outcome, thus establishing the lexical bias effect. In Baars et al.'s Experiment 2, all target items and biasing items were nonwords, but in one list there were real-word fillers (mixed context)[2] while in another list all materials were nonwords (nonlexical context). A lexical bias effect occurred only in the mixed context (36 lexical to 14 nonlexical outcomes in the mixed context and 40 lexical to 32 nonlexical outcomes in the nonlexical context). In particular, nonword outcome items in the mixed context yielded (descriptively) fewer errors than the other three conditions. This finding was taken to support the monitoring account: If words occur in the context, the monitor would employ a lexicality criterion and filter out more nonword errors than word errors. But in a pure nonword context, the monitor would not employ the

[1] Both examples are from Baars et al. (1975, Experiment 1).

[2] The mixed context is often, misleadingly, referred to as the 'lexical context.'

Table 1
Set-up of the SLIP task, with targets used by Baars et al. (1975, Experiment 1)

Item type	Lexical outcome	Nonword outcome	Task
Biasing item 1	*ball dove*	*ball dove*	Read silently
Biasing item 2	*bin dear*	*bin dear*	Read silently
Neutral Filler	*race mice*	*race mice*	Read silently
Biasing item 3	*bark dome*	*bark dome*	Read silently
Target	*darn bore*	*dart board*	Read aloud

lexicality criterion and filter out as many word errors as nonword errors.

However, there are two methodological problems with that experiment. First, Baars et al. (1975) did not test whether there was an interaction between context and outcome. Instead, they performed comparisons between the lexical outcome and nonlexical outcome conditions within each level of context, and this was significant only in the mixed context condition. Possibly, the lack of effect in the nonword context was simply due to a lack of statistical power. Second, Baars et al. did not use counterbalanced lists. A valid comparison across context requires that the same stimuli are tested in each context. Since they tested different items in each context, it is possible that the form of their interaction (suppression of errors on nonword outcome pairs in the lexical context) reflects differential error proneness of different items.

There is also a theoretical problem with Baars et al.'s account, which we will dub the 'standard monitoring account.' The problem is that the monitor is presumed to be adaptive, so that it gives up on the lexicality criterion in a nonword context, but only half-adaptive, because it persists in using the lexicality criterion in the mixed context even though that criterion is dysfunctional *given the task at hand*. The lexicality criterion is dysfunctional in the mixed context, because 67% of the items in that context are nonwords. Therefore, the lexical status is uninformative as to whether an error occurred. If the monitor were truly adaptive, it would have employed an *anti*-lexical criterion (*is this a nonword*?) in the nonword context (if it is a word it is certainly wrong), but no lexicality criterion in the mixed context (if it is a word, it can be right or wrong, and the same holds for nonwords).

Of course, we cannot be sure that the monitor is really adaptive. But self-monitoring is presumed to be a controlled process that flexibly pays selective attention to specific characteristics of speech (Levelt, 1989). This makes sense, because depending on the situation we need to monitor for different aspects of our speech. For example, compare the situations where one addresses a close friend, a business associate, a king or queen, a child, a foreigner, a group of students; when one is interviewed on radio or television, or when one speaks 'fake Swedish' as a party joke, like the Swedish chef character in the 1970s 'Muppet Show.' The last example is particularly informative, because it illustrates that we can produce deviant speech if we want (such as saying nonsense words) and presumably monitor whether it is still deviant. In the case of Baars et al., a truly adaptive monitor confronted with nothing but nonwords would be expected to 'maintain deviancy' in a similar manner: a real-word error would be likely to be suppressed. An anti-lexical criterion such as this would clearly have no effect on nonlexical errors, and importantly, it would not even come into play in a mixed context, where lexicality provides no additional information about the likelihood of an error.

Descriptively, Baars et al.'s (1975) data are inconsistent with the suggestion that lexical errors should be suppressed in a nonword context; rather, it appears as if a mixed context serves to suppress nonlexical errors. However, because of the methodological problems outlined above (Is there an interaction? What form does the interaction take?), we cannot estimate a base error rate from Baars et al.'s data to enable us to determine which errors are actually suppressed relative to which others.

In contrast to Baars et al., Humphreys (2002) found no evidence for a modulation by context (65 lexical to 36 nonlexical outcomes in the mixed context and 61 lexical to 40 nonlexical in the nonlexical context). Humphreys suggested that the discrepancy between the findings was due to a difference in procedure: Her participants had to complete their response within a fixed time, whereas Baars et al.'s description of their method suggests that their participants had ample time to respond. Humphreys speculated that Baars et al.'s participants named items more quickly in the nonword context, so that there was less time for feedback in that context and the lexical bias effect disappeared. This speculation is implausible, because nonwords tend to be named more slowly than words, and studies on reading have provided evidence for the hypothesis that participants set a "time-criterion," so that there is a tendency to name all items in a list equally fast (Lupker, Brown, & Colombo, 1997). In fact, they observed that "In all cases, the more slowly named stimuli were named more quickly in mixed blocks than in pure blocks" (p. 570). It is therefore much more plausible that speakers in the SLIP task also responded *more quickly* in the mixed context than in the nonword context. Nevertheless, we recorded response latencies in one of our experiments, so as to test Humphreys' speculation.

In order to test whether there really is modulation of the bias effect by lexical context, we conducted three experiments. Experiment 1a kept the materials and procedure as similar as possible to those of Baars et al., except that we tested the same targets in each context, allowing us to statistically compare the results from nonlexical and mixed contexts. We recorded naming latencies, in order to test Humphreys' speculation (faster latencies in the nonword condition). Experiment 1b was a control experiment that used the same materials in a perception task; its purpose was to assess whether the results from production can be explained by the perceptual component of the task (in which case the entire debate between feedback and monitoring explanations of lexical bias would obviously become irrelevant). Experiment 2 tested whether the production results generalize to a different, and better controlled, set of materials and a better design.

Experiment 1a

Method

Participants

Thirty-two students at the University of Edinburgh took part in the experiment.

Materials

We constructed 160 pairs of pronounceable, monosyllabic letter strings. This included 20 target pairs. All target pairs consisted of nonwords that either did or did not result in words when spoonerized. Eight of them spoonerized into correctly spelled words (e.g., *liss kong*) and 12 spoonerized into pseudo-homophones (e.g., *vun wice*). The target pairs were the ones used by Baars et al. (1975, Experiment 2), with minor adaptations for British English (for example, the letter string *dawg* is not a pseudo-homophone of *dog* in the UK). There were two versions of each target pair, which were phonologically as similar as possible so that one version would yield a lexical outcome when spoonerized (e.g., *rafe sode*), and the other would yield a nonlexical outcome (e.g., *rabe sofe*). The target pairs are listed in Appendix A.

We also constructed 60 biasing pairs, three for each target, which were designed to prime exchange errors (e.g., *sade roke* for the target *rafe sode*). The initial consonants of the two onsets in each biasing item (e.g., /s/ and /r/ in the example) were always identical to the initial consonants of the to-be-elicited exchange error. Forty of the biasing pairs occurred twice, so that each target was preceded by five biasing pairs. Thirty-three biasing items consisted of nonwords only (e.g., *sade roke*, occurring in both the mixed context and nonlexical context lists) and 67 biasing items occurred in two versions (nonwords only, e.g., *gake mees*, or matched words that shared the initial consonant and the vowel, e.g., *gain meal*).

Additionally, there were 20 control pairs, which were to be named aloud, but were not preceded by any biasing items. These were included to prevent participants from predicting the occurrence of target items. Six control pairs consisted of nonwords only and these occurred in both the mixed context and nonlexical contexts lists. Fourteen control pairs occurred in two versions: In the mixed context lists, they consisted of real words (e.g., *rise mean*) and in the nonlexical context lists they consisted of nonwords, exchanges of the real word version (e.g., *mise rean*).

There were 21 nonlexical filler pairs (e.g., *lep biss*), 14 of which occurred twice, resulting in 35 nonlexical filler items. Additionally, there were 58 filler pairs (17 occurring twice, yielding 75 filler items). These filler items occurred in two versions: A real-word version (e.g., *damp pit*) and a nonword version (their transposition, e.g., *pamp dit*).

List construction and design

Four lists of 250 items each were constructed. Each list consisted of 20 targets, 100 biasing items, 20 control items, and 110 filler items. In two lists (the nonlexical context lists), all materials (targets, biasing items, control items, and fillers) were nonwords. But in two otherwise identical lists (the mixed context lists), 67 biasing items, 14 control items, and 75 filler items were lexical variants of the nonwords in the other two lists. In the mixed context lists 63% of the items presented and 35% of the items that had to be named were lexical.

First, a pseudo-randomly organized master list was constructed under the constraint that target items were separated from each other or from the beginning of the list by at least eight positions. The five biasing items for each target, along with three randomly selected filler items were distributed randomly across these eight positions. The position of the other materials (remaining fillers and control items) was random. From the master list, four counterbalanced experimental lists were derived so that half of the targets were lexical outcome pairs and half were nonlexical outcome pairs and that across the four lists each version of a target occurred twice (one in the lexical outcome and one in the nonlexical outcome). The same counterbalancing was applied to the control items, biasing items, and filler items, so that the four lists either contained identical items or matched items for any particular position (be it target, control, biasing, or filler item). This design improves upon that of Baars et al. (1975, Experiment 2), because these authors used different targets in the two contexts.

Procedure

Participants were tested individually in a quiet room. They wore headphones and throughout the experiment they were exposed to loud, computer-generated noise (following the procedure by Baars et al.). The participants were told to silently read word pairs, but to name aloud the last word pair they had seen whenever a beep sounded. Each word pair was presented for 700 ms and was followed by a blank screen for 200 ms. Targets and control items were followed by a tone which was also presented by the headphones. Tone onset was simultaneous to the offset of each word pair. Naming latencies were recorded using a voice key. In order to encourage participants to speak fast, a second (warning) tone sounded if the voice key was not triggered within 500 ms of the first tone. The next item was presented 950 ms after the voice key had been triggered. This procedure, which gave participants unlimited time to initiate a response, is therefore different from the one employed by Humphreys (2002), but similar to the one employed by Baars et al. (1975). The experiment took about 5 min.

The productions were recorded on Digital Audio Tape, and were scored off-line. The productions were scored as correct (*pote vark* ⇒*pote vark*), full exchanges (*pote vark* ⇒*vote park*), partial exchanges (anticipations, e.g., *pote vark* ⇒*vote vark*and perseverations, e.g., *pote vark* ⇒*pote park*), other errors (*pote vark* ⇒*bote vark*) or a failure to respond (e.g., *don't know*). Aborted errors that rendered the first phoneme of the second word (e.g., *vote p…*) were assigned to the same category as the full response would have been (in the example, full exchange).

Results and discussion

The 640 target trials were correctly named in 524 cases (81.9%) and there were 26 full exchange errors (4.1%), 15 partial exchanges (2.4%), 66 other errors (10.4%), and 9 failures to respond (0.9%). The number of other errors or nonresponses did not differ significantly among the four conditions. In all analyses, we have collapsed the categories of full and partial exchange errors (because there were not enough errors to warrant separate analyses). However, the number of full errors, when considered separately, produced essentially the same pattern. The number of errors in each experimental condition is listed in Table 2.

Most errors occurred for lexical outcome pairs in the mixed context, while there were only few errors in the other three conditions (see Table 2). An initial analysis used the same tests as the ones employed by Baars et al. (1975), namely a nonparametric comparison between the numbers of lexical and nonlexical outcomes for each level of context, by-subjects only. On the Wilcoxon signed ranks test, there was a significant lexical bias effect in the mixed context [6 ties; $Z = 2.24$, $p < .05$] but not in the nonlexical context [9 ties; $Z = .27$, $p = .79$]. We thus replicated Baars et al. (1975) findings: The lexical bias effect only occurred in the mixed context.

An ANOVA on the number of exchange errors with context (mixed vs. nonlexical) as a between-subject and within-item variable and lexical outcome (lexical vs. nonlexical) as a within-subject and within-item variable, showed marginal main effects of context [$F_1(1,30) = 3.13$, $p = .09$; $F_2(1,19) = 4.13$, $p = .06$] and of outcome [$F_1(1,30) = 3.34$, $p = .08$; $F_2(1,19) = 2.17$, $p = .16$]. In the by-subjects analysis, there was a significant Context × Outcome interaction [$F_1(1,30) = 4.70$, $p < .05$; $F_2(1,19) = 1.86$, $p = .19$]. Comparisons between the number of lexical outcomes and nonlexical outcomes for each level of context showed there was a significant lexical bias effect in the mixed context by-subjects [$t_1(15) = 2.53$, $p < .05$; $t_2(19) = 1.55$, $p = .14$] but not in the nonlexical context [$t_1(15) = 0.27$; $t_2(19) = 0.27$]. Comparisons between nonlexical and mixed context for each type of error, showed that lexical errors varied as a function of context [$t_1(30) = 2.18$, $p < .05$; $t_2(19) = 1.82$, $p = .09$], but nonlexical errors did not [$t_1(30) = 0.28$; $t_2(19) = 0.27$].

An analysis was also conducted on the response latencies for correctly produced experimental targets and also on the correctly produced control items (see Table 3). Latencies below 200 ms or above 2000 ms were excluded (2.4%) after which 974 observations (targets and control items) remained. In agreement with a prediction one can derive from Lupker et al. (1997) the targets were read *more slowly* in the nonlexical context than in the lexical context [$F_1(1,30) = 5.70$, $p < .05$; $F_2(1,19) = 57.04$, $p < .001$] and this also held for control items [$t_1(30) = 3.09$, $p < .01$; $t_2(19) = 5.97$, $p < .001$]. Therefore, Humphreys' (2002) suggestion that targets

Table 2
Number of (full and partial) exchanges in each condition of Experiment 1a

Context	Lexical outcome			Nonlexical outcome		
	Full	Partial	Total	Full	Partial	Total
Mixed	10	10	20	6	2	8
Nonlexical	6	0	6	4	3	7

Table 3
Mean response latencies in ms (*M*) and standard errors of the mean (*SE*) for control items, and lexical outcome and nonlexical outcome targets as a function of context (Experiment 1a)

Context	Control items		Lexical outcome		Nonlexical outcome	
	M	*SE*	*M*	*SE*	*M*	*SE*
Mixed	890	36	862	50	919	50
Nonlexical	1047	36	1036	41	1013	27

Note. Data averaged over participants.

are read *more quickly* in the nonlexical context (and that this accounts for the lack of lexical bias in the nonlexical context) can be rejected.

In sum, we established a significant interaction between context and outcome (at least by-subjects), and parametric comparisons within each level of context were in agreement with a nonparametric analysis: The lexical bias effect is restricted to the mixed context. However, it is important to note that the form of the interaction does not support the standard monitoring account. A lexicality criterion that is selectively applied in the mixed context would reduce the number of nonword outcomes in that condition as compared to the nonlexical context and leave the word-outcome errors unchanged. This was the (numerical) pattern observed by Baars et al. (1975). In the present study, however, it is exactly the other way around: the nonword outcomes are unchanged across conditions and the word-outcomes are reduced in the nonlexical context (as opposed to those in the mixed context). Note that this is the pattern predicted by an anti-lexicality criterion that is selectively applied in the nonword context. Since Baars et al. used different items in each context and did not submit their across-context comparisons to statistical tests, we suggest that our reading of the results is more valid. It appears that the suggestive pattern of results Baars et al. observed was an unfortunate by-product of their lack of experimental control.

Before we can attribute the lexical bias and its modulation to any kind of monitoring criterion, however, we need to address a potential caveat that applies to *all* studies using the SLIP task. As the task requires the visual recognition of written words and nonwords, we conducted a control experiment (Experiment 1b) in order to check whether the lexical bias effect is (partly) a result of input processes.

Experiment 1b

Bock (1996) criticized the literature on reading for interpreting the results from word naming studies as being primarily reflective of visual word recognition processes, ignoring the complexities of the production component. We completely agree with her criticism. However, the language production literature, when interpreting data from the SLIP task, is guilty of the converse: The SLIP task requires reading of words and nonwords, but no attention has been paid to any possible role of input processing (but see Camden, Motley, & Baars, 1982). However, there is evidence that readers sometimes substitute a letter from another, simultaneously presented, word (McClelland & Mozer, 1986; Mozer, 1983; Van der Velde, 1992), suggesting that there could be a reading component to the SLIP task. In these reading experiments, two words are briefly visually presented and participants name one of them on a cue. For example, participants see LACE and LINE and have to report LINE. To the extent that participants report LANE more often than, say, LONE, the letter ⟨A⟩ is thought to have migrated from the first word to the second. These migration-errors are more frequent when the outcome forms a real word (McClelland & Mozer, 1986; Van der Velde, 1992), thus constituting a lexical bias. Because this lexical bias also occurs in a task variant that does not require overt production (i.e., a two alternative forced choice task, McClelland & Mozer, Experiment 2) it cannot be (exclusively) due to speech production mechanisms. Indeed, McClelland and Mozer attributed the lexical bias in reading to activation of word-representations in the orthographic lexicon.

Additionally, there are many demonstrations that reading of nonwords also results in activation of lexical representations, in particular of "neighbors" (i.e., real words that differ in only one letter). In lexical decision, it is more difficult to correctly reject a nonword if it has more neighbors (e.g., Andrews, 1989; Carreiras, Perea, & Grainger, 1997), priming studies show that nonword primes facilitate recognition of neighbor targets (e.g., Ferrand & Grainger, 1992, 1993, 1996; Forster & Taft, 1994; Forster & Veres, 1998; Frost, Ahissar, Gotesman, & Tayeb, 2003), and naming studies suggest that nonwords inherit the pronunciations of their word neighbors (e.g., Andrews & Scarratt, 1998; Kay & Marcel, 1981). All these studies suggest that reading a nonword (e.g., POTE) will activate lexical neighbors (including VOTE).

It is conceivable therefore that the lexical bias effect in Experiment 1a resulted from the activation of the lexicon during reading. Reading a lexical-outcome pair such as POTE VARK will activate the word-outcomes VOTE and PARK, but reading the target pair VONE

PANK will not activate the nonword-outcomes PONE and VANK. It is also conceivable that the lexicon becomes especially active in a context containing real words.

To disentangle contributions of perception and production to the errors observed in the SLIP task, we conducted an experiment that was identical to Experiment 1a in its materials and procedure, but with the twist that the target and control items had to be compared to a visually presented probe pair instead of being named: Participants decided whether the probe pair was identical to the last pair or not. For the critical trials, the probes were identical to the targets, so that perception errors would manifest themselves as a miss (i.e., an incorrect NO response). This ensured that participants were not exposed to any real words. Moreover, this allowed both full and partial exchanges to manifest themselves as errors. In contrast to Experiment 1a all control pairs were selected to have nonlexical outcomes when exchanged. The probes on control trials were exchanged versions of the preceding items. In this way, it was ensured that participants would not neglect the small differences between the controls and their probes (and hence would be sensitive to subtle misperceptions of the targets). The hypothesis that the lexical bias effect, or its modulation with context, is due to perceptual processes predicts that a lexical bias (or context modulation) is also apparent in the perception task. But if the effects are genuinely production effects, there should be no effect of lexical outcome, and no outcome by context interaction in perception.

Method

Participants

Twenty-four students at the University of Edinburgh took part. None of them participated in Experiment 1a.

Materials, list construction, and design

The materials and stimulus lists were identical to the ones in Experiment 1a.

Procedure

The procedure was identical to that of Experiment 1a, with one exception. After a target or control item, a probe was presented for 100 ms and the participants were asked to verify as quickly and accurately as possible whether the probe matched or mismatched the last item, by pushing the appropriate button on a button box. Probes for target items were always identical to the target (requiring a 'yes'-response), and probes to control items were always exchanges of the control item (requiring a 'no'-response). Therefore, the correct answer was 'yes' on 50% of the probe-trials, and participants never saw a real-word probe.

Results and discussion

The total numbers of errors (failures to correctly respond 'yes' to veridical probes) are listed in Table 4. Importantly, there was no effect of lexical outcome [$F_1 < 1$; $F_2 < 1$], and there was no Outcome × Context interaction [$F_1 < 1$; $F_2 < 1$]. Unexpectedly, there were more errors in the mixed context than in the nonlexical context [$F_1(1,46) = 7.97$, $p < .01$; $F_2(1,19) = 14.17$, $p < .01$]. When probes were correctly identified, there was a difference in response latencies in that identifications in the mixed context took longer, but this effect was far from significant by subjects [735 vs. 707 ms; $F_1 < 1$; $F_2(1,19) = 15.38$, $p < .01$]. No other effects or interactions were significant.

The main question in this control experiment was whether the lexical bias (and more specifically its interaction with context) would be replicated in a purely perceptual task. As there was no difference between misses on items with a lexical outcome (e.g., POTE VARK) and items with a nonlexical outcome (e.g., PONE VANK), and there was not a trace of an interaction with context, we can safely conclude that this was not the case. Consequently, the lexical bias observed in the production version of this task (Experiment 1a) was not due to perceptual processes.

Somewhat surprisingly, there were more perception errors in the mixed context than in the nonlexical context. We can only offer a post hoc explanation for this unexpected result. We suggest that this difference results from readers paying less attention to sublexical characteristics of the items in the mixed context, relative to the nonword context. In language processing, it makes sense to attend to the highest possible level of integration. The occurrence of words in the mixed context may have automatically guided attention to the word level, so that less attention was paid to sublexical characteristics of the stimuli. Therefore, it became harder to tell apart nonwords that differ in only one letter. We leave further tests of this proposal to reading studies; important for our purposes is that visual word recognition cannot account for the lexical bias effect and its dependence on context in production.

Table 4
Number of error responses (incorrect rejections) in each condition of Experiment 1b

Context	Lexical outcome	Nonlexical outcome
Mixed	50	48
Nonlexical	34	27

In conjunction, Experiments 1a and 1b provide evidence for a modulation of the lexical bias effect by context in production, but not in perception. Our results differ from those of Humphreys (2002), who observed no modulation by context. Our results also differ from those of Baars et al. (1975, Experiment 2) with respect to the form of the interaction. We conducted Experiment 2 in order to test whether the same interaction occurs with materials more similar to those of Humphreys and in a design that tested each participant in each context, making cross-context comparisons even more valid.

Experiment 2

Experiment 2 improved upon Experiment 1a in several ways. First, the materials in Experiment 1a and in Baars et al. (1975) were not controlled for whether the expected spoonerisms were orthographically real words (e.g., when *mave geet* exchanges to *gave meet*) or whether they were pseudo-homophones of real words (e.g., the second word when *fot gude* exchanges to *got fude*). Humphreys (2002) instead only used materials that also form words when the two initial letters were exchanged orthographically. If this difference were relevant one would expect the interaction to disappear. Second, in Experiment 1a context was varied between-subjects. We now opted for a within-subject design, so that any differences between the two contexts cannot be attributed to between-subject variability in error-proneness.

Method

Participants

Forty students at the University of Edinburgh participated. None of them participated in the previous experiments.

Materials

A new set of materials was constructed consisting of four lists of 500 pairs of pronounceable, monosyllabic letter strings. Each participant was presented with two lists (one of which contained 50% real words, and one of which contained only nonwords). Each of the four lists included 250 nonword pairs, comprising 24 target pairs, 120 biasing pairs, 26 control items, and 80 nonword fillers. Additionally, each list contained 250 context fillers (either words or nonwords). All words and nonwords used in the experiment were of the form CVC(C).

There were two versions of each target pair (a lexical outcome and a nonlexical outcome version), which were phonologically as similar as possible (i.e., they shared at least the initial consonant and vowel). The lexical outcome pairs always resulted in orthographically and phonologically real words when the initial letters were swapped. Because we intended to use the same items in an experiment that required a restricted set of phonemes, we further restricted the target pairs to ones with the initial phonemes /t/ and /k/, /s/ and /t/, /s/ and /k/, and /d/ and /g/, and with the vowels /ɜ˞ː/, /ʌ/, /e/, or /ɪ/. Furthermore, we ensured that each word in a given word pair had the same vowel, as repeating the vowel promotes the probability of exchange errors (Dell, 1984, 1986). With the exception of target pairs, no other pair used in the experiment resulted in phonologically or orthographically real words when the initial letters were exchanged. Targets are listed in Appendix B.

The 120 biasing items were made up of 5 nonword pairs for each target. Seventy-two of these pairs shared only the initial consonants with the target, and the remaining 48 shared the vowel as well. The latter 48 biasing items were the only items in the experiment other than the targets to use the target CV onsets.

We also created two sets of 26 control items (to be named by the participant). These were all nonwords. In contrast to Experiment 1a, this allowed for maximal comparability between the mixed and nonlexical conditions by ensuring that participants never produced real words except in error.

Finally, we created 80 fillers, which were all nonwords, and 250 context fillers that occurred in a lexical version and a matched (same initial consonant and vowel) nonlexical version. The context fillers were used to create mixed (50% real words) and nonlexical (all nonword) contexts.

List construction and design

The 24 target pairs were counterbalanced into two sets, consisting of 12 lexical outcome and 12 nonlexical outcome targets. Each set was combined once with the 250 lexical context fillers and once with the 250 nonlexical context fillers, resulting in four lists. To each of these lists we added the 80 nonword fillers, the 120 biasing items, and one of the two sets of 26 control items. Finally, the lists were paired into two experimental treatments, such that each participant would see two lists of 500 items. One of the lists in each treatment contained mixed context items; the other contained the matched nonlexical context items, together with the counterbalanced set of targets (and the other of the set of 26 control items).

Each participant was presented with one combination of two lists. Each list started with a randomized run of 18 (context and nonword) filler items and two of the control items for practice. Following this, each

Table 5
Number of full and partial exchanges in each condition of Experiment 2

Context	Lexical outcome			Nonlexical outcome		
	Full	Partial	Total	Full	Partial	Total
Mixed	21	20	41	4	8	12
Nonlexical	9	15	24	10	7	17

block of 20 items included 10 context fillers, 3 nonword fillers, one control item, 5 biasing items, and a target. The biasing items were assigned pseudo-randomly to the 7 positions preceding the target. To increase the probability of errors, we ensured that the two items directly preceding each target were biasing items, and that these were biasing items that overlapped in both the consonant and the vowel with the to-be-elicited exchange error.

Procedure

This was identical to Experiment 1, with the exception that we now presented two blocks of 500 items (a mixed context block and a nonlexical context block). There was a break in between the blocks. Half the participants received the mixed block first. The experiment took about 10 min.

Results and discussion

The 1920 target trials were correctly named in 1559 cases (81.2%), and there were 44 full exchange errors (2.3%), 50 partial exchanges (2.6%), and 267 other errors (13.9%). The numbers of other errors did not differ significantly between conditions. Since there were now sufficient observations, we report both analyses that collapse full and partial exchanges, as well as analyses restricted to full exchanges. The number of full and partial exchanges in each experimental condition is listed in Table 5.

An ANOVA on the total number of exchange errors (full and partial combined) with context (mixed vs. nonlexical) as a within-subject and within-item variable and outcome (lexical vs. nonlexical) as a within-subject and within-item variable, showed no effect of context [$F_1 < 1$; $F_2(1,47) = 1.17$, $p = .29$], an effect of outcome [$F_1(1,39) = 18.16$, $p < .01$; $F_2(1,47) = 9.26$, $p < .01$] and, importantly, a Context × Outcome interaction [$F_1(1,39) = 4.14$, $p < .05$; $F_2(1,47) = 4.84$, $p < .05$]. Comparisons between the number of lexical outcomes and nonlexical outcomes for each level of context showed a significant lexical bias effect in the mixed context [$t_1(39) = 3.97$, $p < .01$; $t_2(47) = 3.49$, $p < .01$] but not in the nonlexical context [$t_1(1,39) = 1.10$; $t_2(47) = 0.98$].

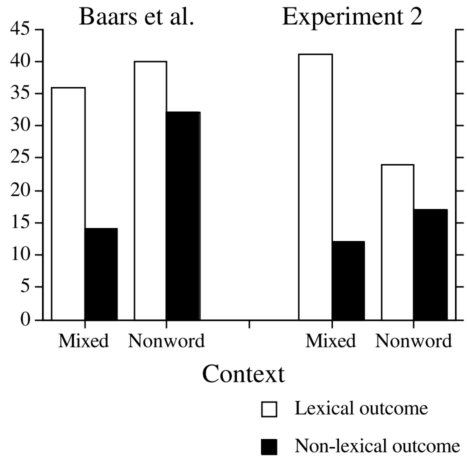

Fig. 1. Number of errors (full and partial exchanges combined) in Baars et al. (1975, Experiment 2) and our Experiment 2.

The same pattern of significance was obtained in the analysis of full exchanges only: no effect of context [$F_1(1,39) = 1.13$, $p = .30$; $F_2 < 1$], an effect of outcome [$F_1(1,39) = 5.03$, $p < .05$; $F_2(1,47) = 5.15$, $p < .05$] and a significant Context × Outcome interaction [$F_1(1,39) = 7.20$, $p < .02$; $F_2(1,47) = 6.19$, $p < .02$]. Comparisons between the number of lexical outcomes and nonlexical outcomes for each level of context showed there was a significant lexical bias effect in the mixed context [$t_1(39) = 3.08$, $p < .01$; $t_2(47) = 3.02$, $p < .01$] but not in the nonlexical context [$t_1(39) = 0.24$; $t_2(47) = 0.23$].[3]

These data firmly establish that the lexical bias effect is modulated by context (in agreement with the conclusions of Baars et al., 1975). However, as in Experiment 1 the interaction took a different form from that observed by Baars et al. To illustrate this, Fig. 1 shows the number of full exchanges in each condition in the Baars et al. experiment and in our experiment.

The Baars et al. data suggest suppression of nonword outcomes in the mixed context, because the mixed context *nonword outcome* condition yielded *fewer errors*

[3] An analysis of partial exchange errors revealed a main effect of outcome [$F_1(1,39) = 8.13$, $p < .01$; $F_2(1,47) = 5.04$, $p < .05$]. The Context × Outcome interaction failed to reach significance.

than the other three conditions (left side of Fig. 1). But our data rather suggest that the mixed context *word outcome* condition yielded *more* errors than the other three conditions (right side of Fig. 1). In order to assess the form of the interaction statistically, we performed Helmert-contrasts. The first contrast compared the mixed context word outcome condition with the mean of the other three conditions. The second contrast compared the mixed context nonword outcome condition with the two nonword context conditions. (The third contrast compared the two nonword context conditions and was therefore identical to the contrast already reported above.)

According to the standard monitoring account, there should be fewer nonlexical errors in the mixed context relative to the three other conditions. In that case, the second Helmert contrast (nonword outcomes in mixed context vs. word and nonword outcomes in nonword context) should be significant. But if there are more word errors in the mixed context than in the other three conditions (as our data suggest), the first contrast will be significant, and this is indeed the case [*All exchanges*: Mixed context, lexical outcome vs. rest: $F_1(1,39) = 8.48$, $p < .01$; $F_2(1,47) = 8.36$, $p < .01$; Mixed context, nonlexical outcome vs. Nonword context: $F_1(1,39) = 2.66$, $p = .11$; $F_2(1,47) = 2.92$, $p = .09$; *Full-exchanges only*: Mixed context, lexical outcome vs. rest; $F_1(1,39) = 7.65$, $p < .01$; $F_2(1,47) = 5.35$, $p < .05$; Mixed context, nonlexical outcome vs. Nonword context: $F_1(1,39) = 3.47$, $p = .07$; $F_2(1,47) = 3.65$, $p = .06$]. In sum, there are significantly more errors in the mixed context, lexical outcome condition than in the other three conditions. But there was no reliable evidence that nonword-errors in the mixed context were suppressed relative to the word- or nonword-errors in the nonlexical context.

General discussion

Our data add to the growing body of evidence that the lexical bias effect is real, in agreement with the seminal studies by Baars and colleagues (Baars et al., 1975) and Dell and colleagues (Dell, 1986; Dell & Reich, 1981) as well as with more recent studies reported by Humphreys (2002) and Nooteboom (2003, in press). Such a further demonstration is important in itself, because some corpus analyses of spontaneous speech errors have failed to find evidence for this effect (Del Viso et al., 1991; Garrett, 1976). We suspect that such failures may be due to the use of suboptimal procedures for estimating chance rate (see Nooteboom, in press, for a new improved method). In any case, the balance of the evidence now clearly supports the existence of a lexical bias effect. Additionally, a control experiment (Experiment 1b) showed that the lexical bias effect (or its interaction with context) cannot be explained by perceptual factors.

Furthermore, the data firmly establish that the lexical bias effect is modulated by context. There is no lexical bias effect if all the items in the context are nonwords, but there is a lexical bias effect if some items are words. As opposed to Baars et al., we tested for an interaction between context and outcome. This interaction was only significant by-subjects in Experiment 1 (a replication of Baars et al.), but it was reliable in Experiment 2, which had more power.

Humphreys (2002, Experiment 4) failed to find an interaction between outcome and context, and she speculated that the reason for the discrepancy between her and Baars et al.'s results is caused by slower naming latencies in the mixed context in the Baars et al. study (where speakers had ample time to respond) but not in her study (where speakers had to meet a very strict response time criterion). Experiment 1a showed that Humphreys' speculation should be rejected: In fact, participants responded *more slowly* in the nonlexical context than in the mixed context, consistent with the proposal by Lupker and colleagues (e.g., Lupker et al., 1997) that participants set a time-criterion as a function of general difficulty. We return to the discrepant findings of Humphreys below.

Importantly, our results differ from Baars et al. (1975) with respect to the *form* of the interaction: In a mixed context, there is no suppression of nonlexical outcomes; instead, there are simply more lexical outcomes in that condition than in the other conditions. Together with Humphreys' (2002) study, these findings cast serious doubt on the standard monitoring explanation of the lexical bias effect. But they also cast doubt on any feedback account *without* a monitor. Notice that feedback and monitoring are sometimes presented as theoretical alternatives, but the real dichotomy is between feedback and feedforward-only theories. There is no reason whatsoever why a feedback-based system would have no self-monitoring of inner speech. In fact, Vigliocco and Hartsuiker (2002) have suggested that the monitor may actually *exploit* feedback, because an aberrant pattern of feedback is a cue that something went wrong (Postma & Kolk, 1993). As we will argue below, the present data support an account that presumes both feedback and the existence of an adaptive monitor.[4]

On an adaptive monitoring account, the monitor sets its criteria flexibly according to task demands. In a context exclusively consisting of nonwords, it will pick up quickly on the nature of the context and set its monitoring criteria

[4] Throughout this paper we have consistently used the term "feedback" to refer to feedback between levels of the production system. However, as Rapp & Goldrick (2004) have pointed out, any model in which later levels of processing affect earlier ones (e.g., Roelofs, 1992, 1997) can be considered to be a feedback model.

accordingly. In such a context, an adaptive criterion is to consider real words as deviations. But in a mixed context, consisting of a mixture of words and nonwords, neither lexicality nor anti-lexicality is a useful criterion. Therefore, in the mixed context the monitor will rely on only one criterion (*is this what I meant to say?*). The pattern of speech errors in that context reflects the number of internal speech errors without monitoring for lexicality: As many lexical errors as nonword errors are covertly repaired. The pattern of speech errors in the nonword context, however, reflects the number of internal speech errors after monitoring for nonlexicality: More lexical errors than nonword errors are covertly repaired.

The data of our experiments are consistent with that hypothesis: if the underlying pattern is more errors with lexical than with nonlexical outcomes (as obtained in the mixed context) then adaptive monitoring will suppress the number of lexical outcomes in the nonlexical context, leading to a reduced, absent, or even reversed lexical bias effect. But this implies that the underlying pattern (before covert repair) already shows lexical bias, and we assume, with Dell (1986), Humphreys (2002) and others that the reason for that underlying bias is feedback.

There is a finding from the ERP-literature that seems consistent with an important aspect of our adaptive monitoring account, namely that if all the items in the context are deviant, the monitor considers a *nondeviant* item as a potential mistake. In particular, Coulson, King, and Kutas (1998) observed that if a grammatical sentence is presented in a large block of ungrammatical sentences, it will elicit a P600-component in the ERP-signal. This component is often thought to reflect reanalysis and repair (Friederici, 2002) or self-monitoring of comprehension (*did I read this correctly?*; Kolk, Chwilla, Van Herten, & Oor, 2003). Coulson et al.'s finding therefore suggests that monitoring comprehension is also adaptive: If everything in the context is ungrammatical, the processes monitoring comprehension reset their criteria (if it is grammatical, it is probably a comprehension mistake).

In contrast to our findings, and apparently in contrast with our adaptive monitoring account, is Humphreys' (2002) report of a lexical bias effect in both contexts. We propose that this is because an important difference in procedure precluded her participants from covertly repairing errors. Her data therefore reflect the situation where there is a feedback, but no monitoring, and this leads to lexical bias in both mixed and nonlexical contexts. The reason that her participants could not covertly repair is simple: They did not have enough time. As mentioned in the introduction, monitoring accounts *must* assume that the monitor has sufficient time to covertly and fluently repair speech errors. But whereas our speakers had ample time to prepare their naming, resulting in naming latencies of around 1000 ms, her speakers had to finish speaking before a 750 ms deadline. This suggests that our speakers began speaking at a moment in time when Humphreys' speakers had already finished. We believe this difference, of roughly half a second, gave our speakers ample time for monitoring and filtering out errors.

Conclusion

In sum, our data show that the lexical bias is really modulated by context, at least when the task circumstances allow the monitor enough time to filter out errors. At the same time, our data strongly argue against both a simple feedback account (there is feedback, but no covert repair) and against a maladaptive monitoring account (the monitor uses a lexicality criterion in the mixed context, but not in the nonword context). Together with the data from Humphreys (2002), they argue for feedback in the language production system, leading to a lexical bias in the underlying pattern of speech errors, before covert repair. If the monitor has sufficient time, it will filter out some of these errors, but according to *smart* criteria: If everything else is a nonword, it is wrong to say a word; but if some items are words and others are nonwords, lexical status is uninformative—so the monitor is not bothered about it.

Appendix A. Target word pairs in Experiment 1 (deviations from Baars et al., 1975; Experiment 2 are indicated with an '*')

Lexical outcome	Nonlexical outcome
rafe sode	rabe sofe
mave geet	mafe geeb
fot gude	fov goom
vun wice	vum wige
wice nin	wibe nid
liss kong	lif kons*
dack pog*	dag pawk
lale peef	lafe peeb
bain med	bape mek
gad boof	gaz boov
pote vark*	pade vard*
kip zote	kib zobe
feep kive	feeb kise
dop tole	dob tov
gize wod*	gike wof*
dood geal	dook geez
set goop[a]	sen goom
doan tef	doak tep
gook tood*	goove toope
guss bon	guz bof

[a] The use of the real word *set* is a mistake that apparently escaped Baars et al.'s attention. To maximize comparability with their study, we have left it in.

Appendix B. Target word pairs in Experiment 2

Lexical outcome	Nonlexical outcome
turb curn	tursh curg
susk tunk	sulm tust
tult cug	tulm cust
seg kell	sep kem
sen telf	selk tem
dulp guck	dulb gups
tid kiff	tig kish
dut gubs	dulk gusp
sult cung	sug culp
tull cuck	tust culm
dush gusk	dusp gulk
dift gim	disp gis
tept kest	tep kes
sut cun	sulp cug
dild gish	dis gisp
sug tulk	sust tulm
siff tink	sim tisp
siss kilk	sint kimp
siln kift	simp kint
tiss kint	tish kig
sest tect	sem telk
dulf gub	dups gulb
surse curf	surn curp
surn turge	sursh turm

References

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 802–814.

Andrews, S., & Scarratt, D. P. (1998). Rule and analogy mechanisms in reading nonwords: Hough dou peaplel rede gnew wirds?. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1052–1086.

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14*, 382–391.

Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review, 3*, 395–421.

Camden, C. T., Motley, M. T., & Baars, B. J. (1982). Cognitive encoding processes: Evidence for a graphemically based short-term memory. *Human Communication Research, 8*, 327–337.

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 857–871.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*, 21–58.

Del Viso, S., Igoa, J. M., & García-Albea, J. E. (1991). Autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research, 20*, 161–185.

Dell, G. S. (1984). Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 222–233.

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

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes, 5*, 313–349.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20*, 611–629.

Ferrand, L., & Grainger, J. (1992). Phonology and orthography in visual word recognition: Evidence from masked nonword priming. *Quarterly Journal of Experimental Psychology A, 45*, 353–372.

Ferrand, L., & Grainger, J. (1993). The time-course of orthographic and phonological code activation in the early phases of visual word recognition. *Bulletin of the Psychonomic Society, 31*, 119–122.

Ferrand, L., & Grainger, J. (1996). List context effects on masked phonological priming in the lexical decision task. *Psychonomic Bulletin & Review, 3*, 515–519.

Forster, K. I., & Taft, M. (1994). Bodies, antibodies, and neighborhood-density effects in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 844–863.

Forster, K. I., & Veres, C. (1998). The prime lexicality effect: Form priming as a function of prime awareness, lexical status, and discrimination difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 498–514.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*, 78–84.

Frost, R., Ahissar, M., Gotesman, R., & Tayeb, S. (2003). Are phonological effects fragile? The effect of luminance and exposure duration on form priming and phonological priming. *Journal of Memory and Language, 48*, 346–378.

Garrett, M. F. (1976). Syntactic processes in sentence production. In R. J. Wales & E. C. T. Walker (Eds.), *New approaches to language mechanisms* (pp. 231–256). Amsterdam: North-Holland.

Hamm, S., Junglas, K., & Bredenkamp, J. (2004). The central executive as a prearticuliatory control device. *Zeitschrift für Psychologie, 212*, 66–75.

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.

Hartsuiker, R. J., Schriefers, H. J., Bock, J. K., & Kikstra, G. (2003). Morphophonological influences on the construction of subject–verb agreement. *Memory & Cognition, 31*, 1316–1326.

Humphreys, K. R. (2002). Lexical bias in speech errors. Unpublished Doctoral dissertation, University of Illinois at Urbana-Champaign.

Kay, J., & Marcel, A. (1981). One process, not two, in reading aloud: Lexical analogies do the work of non-lexical rules. *Quarterly Journal of Experimental Psychology, 33A*, 397–413.

Kolk, H. H. J., Chwilla, D. J., Van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85*, 1–36.

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.

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 570–590.

McClelland, J. L., & Mozer, M. C. (1986). Perceptual interactions in two-word displays: Familiarity and similarity effects. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 18–35.

Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 21*, 578–594.

Mozer, M. C. (1983). Letter migration in word perception. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 531–546.

Nooteboom, S. G. (2003). Lexical bias in phonological speech errors: Phoneme-to-word feedback or output editing. In M. J. Solé, D. Recassens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona 3-9 Augustus 2003* (pp. 2249–2252). Barcelona: Universitat Autónoma de Barcelona.

Nooteboom, S. G. (in press). Listening to oneself: Monitoring speech production. In: R. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech*. Hove, UK: Psychology Press.

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.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*, 460–499.

Rapp, B., & Goldrick, M. (2004). Feedback by any other name is still interactivity: A reply to Roelofs (2004). *Psychological Review, 111*, 573–578.

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. (2004). Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: Comment on Rapp and Goldrick (2000). *Psychological Review, 111*, 561–572.

Van der Velde, F. (1992). Effect of orthographic preactivation on letter migration. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 449–459.

Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin, 128*, 442–472.