Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Speech Communication 47 (2005) 43–58

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech ☆

Sieb G. Nooteboom *

UiL OTS, Utrecht University, Trans 10, 3512 JK UTRECHT, The Netherlands

Received 14 October 2004; received in revised form 24 February 2005; accepted 25 February 2005

Abstract

This paper confirms and exploits the observation that early overt self-interruptions and repairs of phonological speech errors very likely are reactions to inner speech, not to overt speech. In an experiment eliciting word–word and nonword–nonword phonological spoonerisms it is found that self-interruptions and repairs come in two classes, one class of reactions to inner speech, another with reactions to overt speech. It is also found that in inner speech nonword–nonword spoonerisms are more often rejected than word–word spoonerisms. This is mirrored in the set of completed spoonerisms where word–word spoonerisms are more frequent than nonword–nonword ones. This finding supports a classical but controversial explanation of the well-known lexical bias effect from nonwords being rejected more frequently than real words in inner speech. This explanation is further supported by an increasing number of overt rejections of nonword–nonword spoonerisms with phonetic distance between error and target, and increasing lexical bias with phonetic distance. It is concluded that the most likely cause of lexical bias in phonological speech errors is that nonword errors are more often detected, rejected, and repaired than real-word errors in self-monitoring of inner speech.

Keywords: Production; Speech errors; Self-monitoring; Lexical bias

☆ Louis Pols (2004) recently drew attention to the multi-disciplinary, many-sided character of phonetics, already apparent in the topics supposed to be covered at the first International Congress of Phonetic Sciences in 1928, ranging from the physiology of speech and voice, via the development, evolution and anthropology of speech and voice, phonology, linguistic psychology, pathology of speech and voice, comparative physiology of the sounds of animals, even to musicology. In this contribution to a special issue at the occasion of his retirement, I cordially invite Louis to change perspective for a moment, and to allow the possibility that the questions we ask may at times be more uniting than the disciplines and subdisciplines that make up our field. The main question in this paper is "what causes segmental speech errors, other things being equal, to make up real words more often than nonwords?" If it is necessary, in order to answer this question, temporarily to behave like an experimental psychologist, so be it.

* Tel.: +31 302536006/6060; fax: +31 302536000.

E-mail address: sieb.nooteboom@let.uu.nl

0167-6393/$ - see front matter © 2005 Elsevier B.V. All rights reserved.

doi:10.1016/j.specom.2005.02.003

1. Introduction

Speakers every now and then make an error of speech. Speech errors may be phonological, lexical, or grammatical, or they may be errors against social appropriateness. In this paper the focus will be on phonological errors, in particular spoonerisms such as DARN BORE for BARN DOOR. Speakers also monitor their own speech for speech errors, and often detect and repair such errors. There are good reasons to believe that speakers not only monitor their own overt speech, but also their own inner speech, in order to detect and repair errors before they are spoken. Levelt (1989) and Levelt et al. (1999) assume that self-monitoring of inner and overt speech for speech errors is basically the same mechanism, employing the same global, easy-to-use criteria for detecting speech errors. These global criteria are of the form "Is this a word?", "Is this syntactically correct?", "Is this socially appropriate"?

This view of self-monitoring provides a long-standing but still controversial explanation for the phenomenon of lexical bias in phonological speech errors. Lexical bias refers to the fact that phonological speech errors tend to make more real words than nonwords. This has been demonstrated to be a robust phenomenon in laboratory-induced spoonerisms (Baars et al., 1975; Dell, 1986; Motley, 1980; Motley et al., 1982; Humphreys, 2002; Nooteboom, 2003; Hartsuiker et al., 2005a). Lexical bias has, despite some failures to do so (Garrett, 1976; Del Viso et al., 1991), also been convincingly demonstrated in spontaneous speech errors (Dell and Reich, 1981; Nooteboom, 2005). Baars et al. (1975) explained lexical bias from "pre-articulatory editing of inner speech", notably by the assumption that nonword errors are more frequently detected and corrected pre-articulatorily than real-word errors. This explanation presupposes that speech errors in inner speech can be detected, rejected and repaired very rapidly and subconsciously, often leaving no or hardly any trace in overt speech, whereas, of course, repair of speech errors in overt speech is there for all of us to observe. This explanation by Baars et al. of the phenomenon of lexical bias is basically supported by Levelt (1989) and Levelt et al. (1999), who prefer to speak of "self-monitoring" instead of "pre-articulatory editing", and assume that self-monitoring of both inner and overt speech employs the same speech comprehension system that is also used for listening to other-produced speech. For the purpose of self-monitoring, there are supposed to be two forms of input to the speech comprehension system, inner speech, before articulation, closing the inner perceptual loop, and overt speech, via the hearing system, closing the outer perceptual loop. These authors also assume that the self-monitoring system, both in monitoring inner and in monitoring overt speech for phonological speech errors, has no access to the intended word forms, but employs a global criterion of the form "Is this a word?" in detecting speech errors. Detecting and rejecting nonwords more frequently than real words in inner speech would explain the existence of lexical bias.

This perception-based self-monitoring account of lexical bias has not been generally accepted. Lexical bias could potentially also be explained by assuming a production-based monitor (MacKay, 1992; Nickels and Howard, 1995; Postma, 2000; Postma and Oomen, 2005). Both perception-based and production-based self-monitoring would easily account for the fact that lexical bias is sensitive to contextual and situational information, and social appropriateness (Motley, 1980; Motley and Baars, 1979; Motley et al., 1982; Hartsuiker et al., 2005a). However, the major controversy in this area does not seem to be between perception-based and production-based monitoring, but between a monitoring account of lexical bias and a feedback account. Notably those who believe that in the mental preparation of speech there is immediate feedback of activation between phonemes and lexical representations, explain lexical bias (and some other well-attested properties of speech errors) from this feedback (Dell and Reich, 1980; Stemberger, 1985; Dell, 1986; Schade, 1999).

To complicate matters, recently Hartsuiker et al. (2005a) found experimental evidence that led them to assume that the relative frequencies of real-word and nonword speech errors are affected both by Dell-like immediate feedback and by self-monitoring of inner speech. Their basic finding is that in a well-controlled experiment eliciting word–word and nonword–nonword spoonerisms with the so-called SLIP (Spoonerisms of Laboratory-Induced Predisposition) technique (Baars et al., 1975) testing for lexical bias, in which the kind of context is varied from mixed (word–word and nonword–nonword stimulus items) to nonlexical (nonword–nonword items only), it is not the case that nonwords are suppressed in the mixed context, as claimed by Baars et al. (1975), but that lexical errors are suppressed in the nonlexical context. This suppression of real words in the nonlexical context is explained by adaptive behaviour of the self-monitoring system, but this explanation presupposes that there is an underlying pattern, before operation of the self-monitoring system, that already shows lexical bias. This underlying pattern would be caused by immediate feedback as proposed by Dell (1986). An alternative explanation of the Hartsuiker et al. data will be given in the discussion of this paper.

There is a major problem in investigating what the cause is of lexical bias in phonological speech errors. That is that the operation of the proposed mechanisms, either pre-articulatory editing by a self-monitoring system (Baars et al., 1975; Motley and Baars, 1979; Motley, 1980; Motley et al., 1982; Levelt, 1989; Levelt et al., 1999; Hartsuiker et al., 2005a), or immediate feedback of activation within the production system (Dell and Reich, 1980; Stemberger, 1985; Dell, 1986; Schade, 1999), is assumed to be fast, subconscious and apparently often leaving no observable trace in the overt speech. These mechanisms remain largely hidden from observation.

However, there is a possibility that the operation of self-monitoring of inner speech might be made observable. Levelt (1989) has pointed out that in self-repairs like "v...horizontal" detection of the speech error must have been in inner speech, simply because the duration of the "v" sound is shorter than a humanly possible reaction time. In the same vein, Blackmer and Mitton (1991) describe cases where the interval between such early interruptions and following self-repairs is 0 ms, suggesting that in those cases both the self-interruption and the self-repair is planned before the error was made overt. These observations suggest that very early self-interruptions and self-repairs reflect self-monitoring of inner speech, whereas late interruptions and self-repairs would reflect self-monitoring of overt speech. This paper is an attempt to capitalize on the assumption that very early self-interruptions and self-repairs reflect properties of self-monitoring of inner speech, and thus can be exploited to find out whether or not self-monitoring of inner speech may be the cause of lexical bias in phonological speech errors.

The current paper takes its starting point in the view of self-monitoring proposed by Levelt (1989) and Levelt et al. (1999). The following assumptions by these authors are relevant here: (1) self-monitoring for speech errors, among other criteria applies a criterion of lexicality ("is this a word?"), (2) self-monitoring employs the same speech comprehension system that is also used in listening to the speech of others, (3) repairs are made after the monitor has rejected a speech error by re-initiating the production process. It should also be mentioned here that for these authors self-monitoring of inner speech and of overt speech is basically the same mechanism, employing the same criterion. However, recently it has been found that lexical bias in spontaneous speech is not mirrored by overt phonological nonword errors being more frequently repaired than overt phonological real-word errors (Nooteboom, 2005). This either implies that lexical bias is not caused by a criterion of lexicality employed by self-monitoring of inner speech or that self-monitoring employs different criteria for inner and overt speech. Here the focus remains on self-monitoring of inner speech, allowing the possibility that self-monitoring of inner speech has properties that do not show up in self-monitoring of overt speech.

Let us assume, then, that self-monitoring of inner speech employs a criterion of lexicality, detecting and rejecting nonword errors more frequently than real-word errors. Let us also assume, with Levelt (1989; also Postma and Oomen, 2005; Hartsuiker et al., 2005b), that early self-interruptions reflect the operation of self-monitoring of inner speech. We then predict that in an experiment in which nonword errors a priori are equally likely as real-word errors, nonword errors are more frequently early interrupted than real word errors, corresponding with real word errors being more frequently completed than nonword errors. This is the pattern of results one would predict from a self-monitoring account of lexical bias. The experimental technique used in the current investigation for testing this prediction is a well known technique for eliciting spoonerisms, described for the first time by Baars and Motley (1974) and employed by Baars et al. (1975) for studying lexical bias in phonological speech errors. In this technique silent reading of word pairs like DOVE BALL, DEER BACK, DARK BONE, BARN DOOR is used to prime a word–word spoonerism like DARN BORE when the last word pair that has been seen (BARN DOOR) is, after it has disappeared from the visual display, prompted to be spoken aloud. Similarly, word pairs like GIVE BOOK, GO BACK, GAP BOOT, BAD GOOF are used to prime a nonword–nonword spoonerism like GAD BOOF. The current prediction is that (1) cases like DARN BORE instead of BARN DOOR are more frequent than cases like GAD BOOF instead of BAD GOOF, and (2) cases like G...(BAD GOOF) are more frequent than cases like D...(BARN DOOR).

However, we are faced with two questions here. One is: how do we know that a speech error like G...(BAD GOOF) is an interruption of the spoonerism GAD BOOF in inner speech, and not of some other error of speech the lexicality and frequency of which was not controlled for in the experiment? The hard answer is that we don't. However, we will assume here that the relative frequencies of other lexical and nonlexical speech errors than the primed-for spoonerisms also starting with the initial consonant of the second word instead of the first, for example BAD GOOF turning into GAS GOOF, correspond to the relative frequencies of such errors in inner speech. This makes it possible to estimate the relative frequencies of such speech errors in inner speech during the experiment, under the null hypothesis that detection of such speech errors is independent of the lexicality of the speech error. If this null hypothesis holds, the relative frequencies of early interrupted errors should not be significantly different from the estimated relative frequencies. This provides a safeguard against unwarranted conclusions. The other question is: how early must an interruption be in order to reflect a reaction to inner speech instead of to overt speech? Liss (1998) suggested that (at least in spontaneous speech of her apraxic speakers) self-interruptions with error-to-interruption intervals less than 500 ms reflect reactions to inner speech. This seems a somewhat long interval for normal speakers. Also, it has been pointed out that each time value for the criterion separating between reactions to inner and to overt speech on the basis of error-to-interruption interval, is arbitrary (Hartsuiker et al., 2005b). The latter authors employ a probabilistic model for calculating the relative contributions of inner and overt speech to the overall pattern of detection, rejection and repairing of speech errors. Here it will be assumed that if indeed self-interruptions come in two categories, one for reactions to inner and one for reactions to overt speech, error-to-interruption speech fragments will show a bimodal distribution, making it plausible that if we concentrate on the brief speech fragments, these cases will be, at least statistically, dominated by reactions to inner speech.

If, as assumed by Levelt et al. (1999), self-monitoring is perception-based, it stands to reason that the to-be-monitored stream of speech is fed through the same word recognition system that is also used in the perception of other-produced speech. This idea has been worked out by Roelofs (2005), who enriched his model of word production WEAVER++ with a self-monitoring component employing word recognition, and who also presents evidence that phonological words instead of phoneme strings are monitored. Now if indeed, self-monitoring employs the same word recognition system that is employed in the perception of other-produced speech, one may expect the same or similar misperceptions to occur. If the monitor is checking whether a stimulus sound form is a word, it is satisfied when there exists a lexical representation fitting the stimulus sound form. If no such lexical representation is found, this triggers error detection and rejection of the stimulus sound form. In the Baars et al.-like experiment to be described below, either a word–word spoonerism is primed for or a nonword–nonword spoonerism. In the first case, no error will be detected, at least not by the criterion of lexicality supposedly prevailing in inner speech. In the second case, an error is detected, as soon as the first word of the spoonerism is found to have no lexical representation, unless this nonword is misperceived as a real word. This could potentially be any similar real word in the lexicon, but by the very nature of the experiment the most likely candidate is the correct target form, because this form is highly pre-activated by the silent reading part of the task. So, if the nonword error is very similar to the target form, differing only from it in a single phonological feature, it will possibly be misperceived as the target form, and then no error will be detected. The probability of error detection will increase with phonetic distance between nonword error form and the real-word target form. To test this prediction, phonetic distance between the target form and the primed-for error form is an experimental variable in the experiment to be described.

After the monitor has detected a speech error in inner speech, the error may be repaired. An implication of the self-monitoring account of lexical bias is that this can be done so quickly that often no trace of error detection, rejection and repair is found in the overt speech. This is so, because in spontaneous speech lexical bias is found in phonological speech errors that are not accompanied by hesitations, interruptions and repairs (e.g. Nooteboom, 2005). Kolk (1995) has suggested that if at the moment of repair, the word's lemma is still active, it is sufficient to start the compilation of the form once more. Obviously, the assumption here must be that in cases where the repair leaves no observable trace, there is sufficient time for the recompilation after error detection and before articulation. Whether this is a realistic assumption in the context of the present experiment, is unclear. The time pressure in the experiment may be such that fluent repairs are rare. Here we will not be concerned with fluent repairs, but with overt repairs of errors in inner speech. Levelt (1989) has estimated that even with minimal buffering of inner speech, the monitor has in many cases sufficient time to send an interrupt signal to the Articulator. Of course, when the interrupt-to-repair interval is (close to) 0 ms, as in GBAD-GOOF, the repair must have been planned before speech was interrupted. Such cases, observed by Blackmer and Mitton (1991), make it plausible that in more relaxed conditions, possibly with a greater buffer of prepared inner speech, the repair could have been executed before articulation of the error had started. This implies that the command for interruption and the command for initiating a repair can be issued in parallel.

This paper sets out to confirm and exploit the possibility that early self-interruptions and self-repairs of phonological speech errors reflect self-monitoring of inner speech and not of overt speech. It will be attempted to answer the following questions:

(1) Do overt interruptions and self-repairs clearly fall into two classes, one class to be assigned to self-monitoring of inner speech and one to self-monitoring of overt speech, as suggested by the observations by Levelt (1989) and Blackmer and Mitton (1991)?
(2) If so, do early overt interruptions, possibly reflecting self-monitoring of inner speech, show the effects of a criterion of lexicality applied by the self-monitoring system, in that primed-for nonword–nonword errors are more often early interrupted than primed-for word–word errors.
(3) Is self-monitoring of inner speech (contrary to what has been found for self-monitoring of overt speech; see Nooteboom, 2005), sensitive to the phonetic distance between error and target, in that nonword–nonword errors are more often early interrupted with increasing phonetic distance between error and target? This is what one would expect from a perception-based self-monitoring system, and would explain that lexical bias in phonological speech errors is sensitive to phonetic distance between error and target (Nooteboom, 2005), and thus lend further support to a self-monitoring account of lexical bias.

Below an attempt will be described to answer these questions experimentally, employing the SLIP technique invented by Baars and Motley, 1974, and used by Baars et al. (1975) to study lexical bias in phonological speech errors. The technique was since used many times to study aspects of the mental preparation of speech, and recently has seen a revival.

2. Method

The method used was basically the same as the one applied by Baars et al. (1975): Subjects were to read silently Dutch equivalents of word pairs like DOVE BALL, DEER BACK, DARK BONE, BARN DOOR, presented one by one, until a prompt told them to speak aloud the last word pair seen. Some modifications were made. The essential modifications were meant (1) to make it possible to measure response times (in this paper response times will not be reported), (2) to increase the number of overt repairs, by allowing time for correction, and (3) to demonstrate the effect of phonetic distance between the to-be-spoonerised consonants.

2.1. Stimuli

Stimulus word pairs, to be read aloud, consisted of monosyllabic Dutch CVC words (with a few CVCC exceptions when the language ran out of monosyllables with the required properties), visually presented in clear capital print on a computer screen and intended to be read silently. Each of the 36 test stimulus word pairs was preceded by 2, 3, 4, 5 or 6, precursor word pairs, the average number being 4, the total number being 144. In the priming condition the last 2, 3 or 4 of these precursor word pairs were phonologically priming a word–word or a nonword–nonword spoonerism. On average there were 3 priming word pairs, chosen to prime a spoonerism, as in the sequence "GIVE BOOK, GO BACK, GAP BOOT" preceding the test stimulus "BAD GOOF", with a total of 108. The other precursor word pairs were intended to hide as much as possible the goal of the experiment from the subjects. In the non-priming or base-line condition, there were also 2, 3, 4, 5, or 6 precursor word pairs, on average 4, and in total again 144, non of which of course were priming a spoonerism.

The initial consonants of priming word pairs and test word pairs were chosen from the set /f, s, X, v, z, b, d, p, t, k/. There were 18 test stimuli primed for nonword–nonword spoonerisms, as "BAD GOOF" giving "GAD BOOF", and 18 test stimuli primed for word–word spoonerisms as "BARN DOOR" giving "DARN BORE". Each set of 18 was divided in three groups of six stimuli with equal phonetic distance between initial consonants, viz. 1, 2 or 3 distinctive features. For example, /f/ and /s/ differ in 1 feature, /f/ and /t/ differ in 2 features, and /f/ and /d/ differ in 3 features.

Base-line stimuli were not controlled for expected outcomes of spoonerisms, class of initial consonants, or phonetic distance between target and potential error. In all other respects they were similar to the test stimuli: words used in the test and base-line word pair stimuli were, with a few exceptions, all simple and common monosyllabic words. Low frequency words were avoided. The base-line stimuli were unfortunately not identical to the test stimuli, in which case the only difference would have been in the presence or absence of phonological priming. The reason is that, because in experiments of this nature intersubject variability is enormous, it was deemed important to make intra-subject comparisons.

After each test and each base-line stimulus word pair the subject saw on the screen a prompt SPREEK UIT (= "SPEAK"; for this way of prompting the subject to speak the last word pair seen, see Baars, 1980). After that the subject saw a second prompt CORRECTIE (= "CORRECTION"). In addition to the set of test and base-line stimuli described so far, there was a set of seven stimuli with a variable number, on average 4, and with a total of 28, of non-priming precursor word pairs to be used as practice for the subjects, and, of course, also followed by two prompts each. The total number of visually presented precursor word pairs was 144 for the primed stimuli, plus 144 for the base-line stimuli, plus 28 for the practice stimuli, equals 316. The total number of test stimuli was 36 for the primed stimuli, plus 36 for the base-line stimuli, plus 7 for the practice stimuli, equals 79. The total number of prompts was 79 times the "SPEAK" prompt plus 79 times the "CORRECTION"prompt, equals 158. The total number of visual presentations on the screen was 316 + 79 + 158 = 553.

2.2. Subjects

There were 50 subjects, 17 male and 33 female, all of them naive as to the purpose of the experiment. They were staff members and students of Utrecht University, all with standard Dutch as their mother tongue and with no known history of speech or hearing pathology. Subjects varied in age from 17 to 56.

2.3. Procedure

Each subject was tested individually in a sound proof booth. The timing of visual presentation on a computer screen was computer controlled. The order in which test and base-line stimuli, along with their priming or non-priming preceding word pairs were presented was randomized and different for each subject. Each (non-)priming word pair, each test or base-line stimulus word pair, each SPEAK-prompt, and each CORRECTION-prompt was visible during 900 ms and was followed by 100 ms with a blank screen. The subject was instructed, on seeing the appropriate prompt to speak aloud the last word pair presented before this prompt. The subject was instructed to correct the spoken word pair in case of error. It was not necessary to wait for the "CORRECTION" prompt. The purpose of the latter was only to provide each subject with plenty of time for correction in case an error was made. Testing of each subject took 9 min and 20 s, including the seven practice stimulus pairs. All speech of each subject was recorded, in most cases with a Sennheiser ME 50 microphone and in case of two subjects with a Beyerdynamic DT292 headset, and digitally stored on one of two tracks of DAT with a Grundig DAT-9009 FineArts DAT-recorder with a sampling frequency of 48,000 Hz or, in some cases, 32,000 Hz. The resulting speech was virtually always loud and clear. On the other track of the DAT two tones of 1000 Hz and 50 ms duration were recorded with each test or base-line stimulus, one starting at the onset of the visual presentation of the "SPEAK" prompt, the other starting at the onset of the presentation of the "CORRECTION" stimulus. These signals were helpful for orientation in the visual oscillographic analysis of the speech signals (and also for measuring response times). Whereas Baars et al. (1975) had their subjects listen to white noise during the experiment, probably to make them focus on inner speech instead of on overt speech, this was avoided in the current experiment.

2.4. Collecting the data

After the experiment it turned out that there had been a slight error in the controlling program during part of the experiment. For the stimuli with expected word–word spoonerisms nine subjects had been presented with 17 instead of 18 stimuli. The missing stimulus was different for each of these subjects. Similarly, for the stimuli with expected nonword–nonword spoonerisms, eight stimulus presentations were missing, again a different one for each of eight subjects. From the base-line stimuli eight were missing.

Reactions to all remaining 1783 test stimulus presentations and 1792 base-line stimulus presentations were transcribed either in orthography, or, where necessary, in phonetic transcription by two phonetically trained transcribers, viz. the present author and one of his students. The two transcribers worked independently using a computer program for the visual oscillographic display and auditory playback of audio signals. Afterwards the present author compared the two transcriptions. These differed in less than 2% of all utterances and in less than 10% of all utterances containing an error. In all cases where the transcriptions differed, the original recording was accessed again, and carefully and repeatedly inspected in order to resolve the ambiguity. Whenever an utterance remained ambiguous, the solution closest to the correct form was chosen. This ambiguity nearly always concerned the voiced or voiceless character of the initial consonant. Because of this a few potential speech errors were missed. All intruding sounds, often clearly not response initiations but signs of hesitation such as hmmm, huhh, ghghgh, or blowing-sounds, were transcribed, and later interpreted as hesitations.

3. Results

3.1. Early self-interruptions and self-repairs as reactions to inner speech

The design of the experiment was chosen such that it would make it possible to investigate lexical bias in the primed-for spoonerisms. Such primed-for spoonerisms were, as in other such experiments, relatively rare. There were only 56 (3.1% of all responses to test stimuli) completed spoonerisms of the expected kind, and 67 (3.7%) interrupted speech errors that could have been spoonerisms of the expected kind. I will come back to those separately later. Fortunately the experiment generated somewhat more, viz. 381, speech errors of all possible kinds, such as, apart from the primed-for spoonerisms, not primed-for exchanges, anticipations and perseverations, word blends, intrusions, word salad, interruptions following a correct fragment of speech, hesitations, no responses. Table 1 provides a breakdown of all correct and incorrect responses.

It is noteworthy that there are relatively many (80) intrusion errors. These nearly always were words that had occurred earlier in the experiment, often among the precursor words immediately preceding the stimulus word pair in question. There were no cases of apparent interference with the prompts to speak or to correct and no cases of interference with the next set of precursors. It may also be of interest that there are not less than 20 anticipations of single features, as in SONG BIG turning into FONG BIG. There were also 10 speech errors of a somewhat unexpected kind, where clearly two phonemes were competing to be articulated, as in FKF…KOOP FILM. Such errors suggest that the error form and the target form occasionally are simultaneously present in inner speech. If so, in those cases there would be no need for repair through recompilation of the target form.

Of the 381 speech errors in Table 1, 159 errors were followed by a self-repair. The latter make it possible to see whether or not self-interruptions followed by self-repairs come in two classes. The set of 159 errors was broken down according to the number of phonemes spoken before interruption. The ensuing histogram is given in Fig. 1.

Except the speech errors in the rightmost column, all speech errors in Fig. 1 were interrupted before completion. Most interrupted speech errors are interrupted early, immediately after the initial consonant or else after the initial CV. Other interruptions are rare. If we neglect these very rare cases, speech errors are either early interrupted or completed, suggesting the possibility that early interruptions are reactions to errors detected in inner speech, and that late interruptions are reactions to overt speech. If this is indeed the case, then one would also expect a bimodal distribution of the interruption-to-repair intervals. This is so,

Table 1
Numbers of speech errors of different types

Types of responses	Example	Lexical	Nonlexical
Completed spoonerisms	zeur giek > geur ziek	37	19
Interrupted speech errors	bek som > s…(bek som) kaap vang > vaa…	28	39
Other errors starting with second C1	koop film > fil…koop	21	27
Single feature anticipations	song big > fong big	8	12
Competing sounds	koop film > f.k.f…koop film paf kies > f.k.pfafkies	7	3
Various other sound errors	pit gok > pik gok	11	21
Intrusions	bek som > poel sam	26	54
Hesitations and repetitions	bal pot > uhh bal pot vul bel > vul…vul bel	4	4
No response	soep puf > …	28	22
Total		170	201

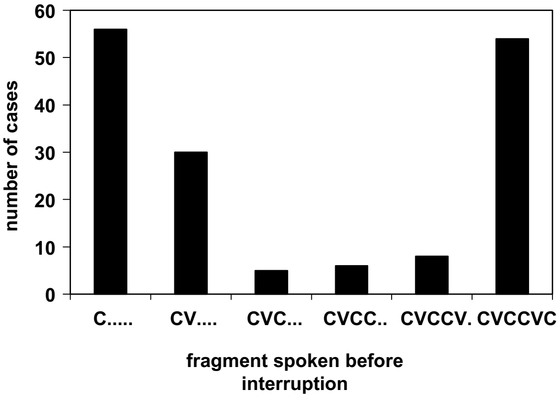

Fig. 1. Breakdown of the set of laboratory-induced speech errors that were followed by a self-repair, according to the number of phonemes spoken. The target always consisted of a CVC–CVC word pair.

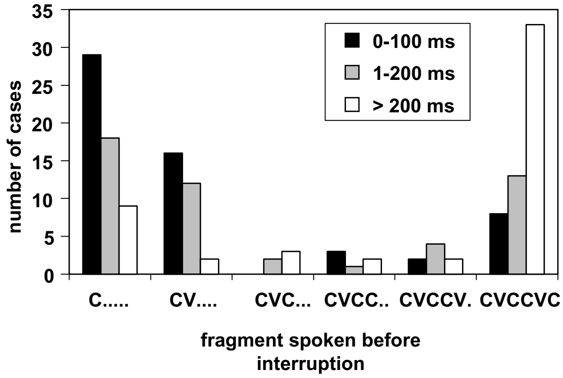

Fig. 2. The same data as in Fig. 1, but now each subset corresponding to a certain length of the spoken fragment is broken down in three classes of duration of the interruption-to-repair interval.

because when an error is detected in inner speech, very likely, at least under assumption that the command for interruption and the command for initiating the repair can be issued in parallel, the repair is also initiated in inner speech, before the interruption command to the Articulator has become effective. Of course, when an error is detected in overt speech, the repair can only be initiated after the speech error has become observable. A bimodal distribution of interruption-to-repair time intervals, with early interruptions followed by short and late interruptions followed by long intervals, would be a strong argument in favour of early interruptions being reactions to inner speech. The relevant data are given in Fig. 2.

The data in Fig. 2 clearly show that early interruptions tend to be followed by very short interruption-to-repair intervals, and late interruptions tend to be followed by much longer interruption-to-repair intervals. In fact, of the 44 cases where the interruptions after only C or CV were followed by interruption-to-repair intervals less than 100 ms, 20 had intervals of 0 ms duration, not counting the six cases where the interruption-to-repair interval coincided with a normal-duration silent interval for a stop consonant. Interruption-to-repair intervals of 0 ms appear to be quite normal, at least following early interruptions. In those cases not only the interruptions but also the repairs must have been planned before overt speech started. They can only have been planned as reactions to inner speech. These data strongly suggest that self-interruptions and self-repairs come in two classes, early interruptions and self-repairs being reactions to errors in inner speech, late interruptions and self-repairs being reactions to errors in overt speech. This answers the first question, and opens the perspective that the normally hidden operation of self-monitoring of inner speech can be caught red-handed in early interruptions and self-repairs.

3.2. Does error detection in inner speech employ a global criterion of lexicality that could explain the lexical bias effect?

We can now turn to the second question, asking whether self-monitoring of inner speech does, and self-monitoring of overt speech does not apply a global criterion of lexicality. To this end, we should concentrate on the primed-for spoonerisms only, separating errors that are detected and interrupted early from errors that are detected late, after completion of the error. In the experimental data there is only one way to decide that a completed spoonerism has been detected by the speaker. That is, when the spoonerism is repaired. In this experiment, despite the extra measures taken to promote repairs, completed spoonerisms are hardly ever repaired. These were overtly repaired only 5 times, three word–word and two nonword–nonword spoonerisms. This is far below the 50% overt repair rate found in spontaneous speech (cf. Nooteboom, 1980, 2005). Interrupted speech errors, like S...BEK SOM, were nearly always repaired. Obviously, subjects stopped speaking in order to make a repair. The number of speech errors interrupted before completion relative to the number of completed spoonerisms (67 compared to 56 = 54%), was not systematically greater than in other similar experiments (Dell, 1986, 68%; Dell, 1990, 32%; Humphreys, 2002, 49%) and obviously varies a great deal from experiment to experiment. The present data seem to make it possible to find out (a) whether in self-monitoring of inner speech a criterion of lexicality is applied, and (b) whether completed spoonerisms (repaired or not) show the well-known lexical bias effect, that possibly could be explained from self-monitoring of inner speech applying a criterion of lexicality. The relevant data are presented in Table 2.

The completed spoonerisms in Table 2 correspond exactly to the primed for spoonerisms: they include errors like GAD BOOF for BAD GOOF, but not GAP BOOF for BAD GOOF. Thus partial spoonerisms are not counted. The interrupted speech errors in Table 2 are exactly equal to the primed-for spoonerisms up to the point of interruption. From the hypothesis that lexical bias is caused by nonlexical errors being more frequently detected in inner speech than lexical ones, the prediction follows that there are more word–word than nonword–nonword completed errors and more nonword–nonword than word–word early interrupted errors, leading to a significantly different distribution of the numbers of word–word and nonword–nonword errors over completed versus early interrupted. This is precisely what Table 2 shows. The word–word spoonerisms in Table 2 were all full spoonerisms with only the two initial consonants interchanged. These data would support a self-monitoring account of lexical bias, if we could be certain that cases like S...BEK SOM in inner speech indeed were also full spoonerisms of the primed-for kind, interrupted before completion. Of course, this is impossible to know. The speech error in inner speech might have been any kind of speech error beginning with an S. Inspection of all errors elicited in the experiment, not represented in Table 2, revealed that stimuli priming for word–word spoonerisms elicited 21 irregular speech errors of very different kinds starting with the initial consonant of the second target word. Of these the first word, which presumably would trigger detection in case the error would have been detected, was a real word in 18 cases, and a nonword in three cases. Similarly, stimuli priming for nonword–nonword spoonerisms elicited 27 speech errors of various kinds starting with the initial consonant of the second target word. Of these 19 started with a real word, and eight with a nonword. Obviously, as soon as a speech error deviates from the primed-for kind, lexical bias gets very strong. Very likely this is also the case in inner speech. So potentially, the interrupted errors considered nonlexical in Table 2 might have been lexical after all. On the other hand, in the condition priming for nonword–nonword errors, there are bound to be more nonlexical errors in inner speech than in the condition priming for word–word errors, and this would explain the difference in distribution between lexical and nonlexical errors in Table 2, because these nonlexical errors would and the lexical errors would not be caught by a criterion of lexicality. Let us assume as a null hypothesis that error detection in inner speech is independent of the lexicality of the primed-for error, and that the relative frequencies of errors starting with the initial consonant of the second word is the same in overt speech and inner speech. We then find a relative frequency of such errors in case of the stimuli priming for word–word errors of 37 word–word errors plus 21 other errors equals 58, giving 0.56 as a fraction of all errors concerned, and a relative frequency of such errors in case of stimuli priming for nonword–nonword errors of 19 nonword–nonword errors plus 27 other errors equals 46, giving 0.44 as a fraction. If indeed the

Table 2
Numbers of laboratory-induced spoonerisms as a function of lexical versus nonlexical outcomes and of completed versus early interrupted ($\chi^2 = 7.21$; df = 1; $p < 0.01$)

	Completed	Interrupted
Lexical	37	28
Nonlexical	19	39

detection of speech errors in inner speech is independent of lexicality of the primed-for error, there would be no significant difference between the actual numbers of interrupted speech errors, 28 for stimuli primed for word–word errors and 39 for stimuli primed for nonword–nonword errors, and the expected numbers based on these fractions, i.e. 37.52 and 29.48 respectively. The relevant data are given in Table 3.

These data show a significant difference between the observed and the expected numbers, in that nonword–nonword errors are more frequently and word–word errors less frequently interrupted than expected under the assumption that error detection in inner speech is independent of the lexicality of the primed-for spoonerism. This provides support for the self-monitoring account of lexical bias. It should be noted, though, that interrupted word–word errors do occur. Inspection reveals that error-to-interruption fragments and interruption-to-repair intervals (the data shown in Figs. 1 and 2) show the same distribution for word–word errors as for nonword–nonword errors. Apparently, word–word errors can be detected by self-monitoring in inner speech, and equally fast as nonword–nonword errors, albeit less frequently. This remains unaccounted for in Levelt's self-monitoring theory.

3.3. Does the detection rate of nonword errors depend on phonetic distance between error and target?

If self-monitoring of inner speech applies a criterion of lexicality and is perception-based, and

Table 3
Observed and expected numbers of laboratory-induced early interrupted speech errors as a function of the lexicality of the primed-for spoonerism

	Observed	Expected
Lexical	28	37.52
Nonlexical	39	29.48

Expected numbers are based on an estimate of the relative frequencies in inner speech of speech errors starting with the initial consonant of the second target word. The estimate is based on the relative frequencies of such errors in overt speech, within this experiment ($\chi^2 = 5.49$; df = 1; $p < 0.0191$).

also lexical bias results from self-monitoring of inner speech, as proposed by Levelt (1989) and Levelt et al. (1999), one would expect (1) that, for nonwords, the probability of detecting an error in inner speech increases with increasing phonetic distance between error and target, and (2) that, correspondingly, the lexical bias effect increases with increasing phonetic distance. The reason for the first expectation is that the more similar a nonword error is to the target word, the more probable it is that the nonword, for which there is no lexical representation, is recognized as the target word (see Section 1). Within a strict interpretation of the perception-based self-monitoring assumed by Levelt (1989) and Levelt et al. (1999) there is no reason to have a similar expectation for the early interrupted word–word errors, because these would escape the criterion of lexicality. The very occurrence of such early interrupted word–word errors requires a separate explanation. The relevant data are given in Table 4.

The data in Table 4 show that in inner speech detecting real-word phonological errors does not clearly depend on phonetic distance, but detecting nonword phonological errors does. As so far no mechanism has been proposed for detecting real-word errors in inner speech, we had no prediction for the effect of phonetic distance on the probability of detecting real-word errors. The effect of phonetic distance on the probability that a nonword error is detected, is as expected from the hypothesis that self-monitoring employs the word recognition system for testing the lexicality of the phonological word forms in inner speech, as explained in the introduction. Now if, as believed by Levelt (1989) and Levelt et al. (1999), lexical bias is caused by self-monitoring of inner speech rejecting nonwords more frequently than real words, lexical bias should accordingly increase with phonetic distance. This prediction is borne out by the data: There is a significant interaction in numbers of completed spoonerisms between lexicality of the error and phonetic distance in the predicted direction. This confirms a finding in earlier research on spontaneous speech errors, where it was also found that lexical bias increases with increasing phonetic

Table 4
Numbers of speech errors as a function of completed spoonerisms versus early interrupted cases and of phonetic distance in number of features between error and target, separately for (a) word–word spoonerisms ($\chi^2 = 3.31$; df = 2; $p > 0.1$; not significant) and (b) nonword–nonword spoonerisms ($\chi^2 = 9.51$; df = 2; $p < 0.01$; significant)

	Completed	Early interrupted
(a) Word–word spoonerisms		
1 feature	10	9
2 features	21	9
3 features	6	8
(b) Nonword–nonword spoonerisms		
1 feature	12	11
2 features	6	12
3 features	1	16

There is also a significant interaction between completed word–word and completed nonword–nonword spoonerisms ($\chi^2 = 7.02$; df = 2; $p < 0.03$).

distance between error and target (Nooteboom, 2005). It is unclear, though, why overall there are more spoonerisms with a phonetic distance of 2 than with a phonetic distance of 1 feature. One would have expected the overall numbers of spoonerisms to decrease with increasing phonetic distance, as it is well known that the probability of a phoneme substitution increases with similarity between target and error phoneme (Nooteboom et al., 1969; Fromkin, 1973; MacKay, 1973). Given that the somewhat high number of spoonerisms with a phonetic distance of 2 features is due to a single cell (completed lexical errors), it may be assumed that this reflects noise in the relatively scarce data. The answer to the second question, then, is that, as predicted from a perception-based self-monitoring account of lexical bias, and from the fact that lexical bias is sensitive to phonetic distance between error and target, self-monitoring of inner speech as made observable in early interrupted spoonerisms, is sensitive to phonetic distance between error and target in detecting nonword errors.

Summarizing: results have been obtained demonstrating that:

(1) laboratory-induced repaired spoonerisms come in two classes, one class of early interrupted spoonerisms, supposedly being reactions to inner speech, and one class of completed spoonerisms, repairs of which clearly are reactions to overt speech,
(2) nonword–nonword spoonerisms are more frequently interrupted at an early stage than word–word spoonerisms, and there are more completed word–word spoonerisms than completed nonword–nonword spoonerisms,
(3) the number of early interrupted nonword–nonword spoonerisms increases and the number of completed nonword–nonword spoonerisms decreases with phonetic distance between error and target, whereas the number of early interrupted word–word spoonerisms shows no clear effect of phonetic distance between error and target, and the number of completed word–word spoonerisms shows no interpretable pattern as a function of phonetic distance.

4. Discussion

The finding that self-interruptions and self-repairs so clearly fall in two separate classes, one class in which planning the interruption and the repair must have taken place before the error was made overt, and the other where planning interruption and repair most likely have taken place after the error was detected in overt speech, supports Levelt's dual perceptual loop theory of self-monitoring (Levelt, 1989; Levelt et al., 1999), by providing evidence for the idea that there are two separate stages during which speech errors may be detected, one stage before speaking, the other after speaking. There is little doubt that self-monitoring for speech errors in overt speech is done by listening to one's own audible voice: Every now and then we consciously hear our own overt speech errors. Moreover, masking the speaker's own speech with white noise effectively decreases the rate of error detection and repair (Postma and Kolk, 1992).

The approach taken in this paper leans on the assumption that early self-interruptions and early self-repairs reflect the operation of self-monitoring for speech errors in inner speech. The validity of this assumption seems evident, given that the speakers would have had insufficient time to plan their interruptions and repairs on the basis of the very brief fragments of overt speech (cf. Levelt, 1989). This is supported by the many cases in which not only the interruption is very rapid, but there is also no time lapse between interruption and repair. Inevitably, in those cases both the interruption and the repair must have been planned before speech became overt (Blackmer and Mitton, 1991). Of course, if both inner and overt speech are to be monitored, the speaker's attention must be divided or alternated between inner and overt speech. This may explain why the percentages of interrupted spoonerisms vary between different experiments (Dell, 1986, 68%; Dell, 1990, 32%; Humphreys, 2002, 49%; the current experiment 54%). Differences in experimental set-up, procedure, and instruction may easily influence the division of attention. The "division of labour" between monitoring inner and monitoring overt speech was recently captured in a probabilistic model, that nicely accounts for existing data on speech error detection rates in normal speech, noise-masked speech and speech in Broca's aphasics, under the assumption that this division of labour is under the control of selective attention (Hartsuiker et al., 2005b).

The main finding in the present paper is that nonword–nonword spoonerisms are more frequently early interrupted than word–word spoonerisms, mirroring the lexical bias found in completed spoonerisms. Assuming that, as argued above, early interruptions indeed reflect the operation of self-monitoring of inner speech, this finding provides evidence in favour of a self-monitoring account of lexical bias. This does not exclude the possibility that lexical bias in phonological speech errors has some other additional source, such as production-based monitoring (MacKay, 1992; Nickels and Howard, 1995; Postma, 2000; Postma and Oomen, 2005), or immediate feedback of activation from phonemes to lexical representations as proposed by Stemberger (1985), and Dell (1986). Schades model (1999) simulates some monitor-like behaviour with the help of a mechanism that is analogous to the binding and checking in WEAVER++ (Levelt et al., 1999), and behaves as if it applies a criterion of the form "is this a word?" (Schade, 1990). This is, of course, in direct competition with Levelt's account of lexical bias. It should be noted, though, that Schade's model does not account for context effects on lexical bias as reported by Motley et al. (1982) and by Hartsuiker et al. (2005a). Such context effects remain an argument in favour of a self-monitoring account of lexical bias.

Schade's model is also interesting in another respect. The lexical bias effect in this model is weak. This would be congruent with the suggestion by Hartsuiker et al. (2005b) that there are two sources of lexical bias, one being the Dell-like immediate feedback of activation, the other being self-monitoring, as assumed by Hartsuiker et al. (2005b). The reader will recall that these latter authors found suppression of real words in a nonlexical context, the result being that in the nonlexical context both spoonerisms with lexical outcomes and spoonerisms with nonlexical outcomes are frequently rejected. They explained this from adaptive behaviour of the self-monitoring system, but also stated that this explanation presupposes that there is an underlying pattern, before operation of the self-monitoring system, that already shows lexical bias. This underlying pattern would then be caused by immediate feedback as proposed by Dell (1986). Despite this congruence between Schade's model and the suggestion by Hartsuiker et al. (2005a), it seems, given the present data, more parsimonious to explain the data obtained by Hartsuiker et al. (2005a) by assuming that the self-monitoring system in a context having lexical stimuli (whether or not mixed with nonlexical ones) would behave as it would in everyday life, and reject nonwords more easily than words, whereas in a quite abnormal context with nonlexical stimuli only, the self-monitoring system would tune itself to this context and reject real word errors more easily than it otherwise would, but also cannot stop itself from the habitual rejection of nonword errors. Likewise, although there is no definitive proof that lexical bias is caused by self-monitoring of inner speech rejecting nonwords more easily than real words, it seems in the context of the present findings most parsimonious to make that assumption. Of course, those who believe in immediate feedback of activation, might claim that feedback models are more parsimonious in explaining phenomena such as lexical bias than models without feedback but enriched with a self-monitoring component. But the evidence for self-monitoring of both overt and inner speech is overwhelming, so a self-monitoring component is needed anyway.

The assumption that lexical bias is caused by lexically sensitive self-monitoring of inner speech is further supported by the finding that the rejection rate for nonword errors in inner speech rapidly increases with increasing phonetic distance between error and target. This might be explained from assuming that in the perceptual system employed by self-monitoring, word forms in inner speech are fed to the word recognition system. When no fitting lexical representation is found, an error is detected and a repair is initiated. However, when the error form is phonetically very similar to the target form it is not unlikely that this target form is recognized, because in the SLIP task this target form is in most cases pre-activated by the silent reading part of the task. The probability that the target form will be recognized on the basis of the error form rapidly decreases with increasing phonetic distance between error and target.

It is more difficult to explain from Levelt's self-monitoring theory that not only nonword–nonword errors but also word–word errors are often early interrupted. These should safely pass the application of a criterion of lexicality. Detection of real-word phonological errors in normal speech is, within Levelt's theory, in principle possible by assuming that the monitor treats these errors as lexical and not as phonological errors, and checks whether the error form is syntactically correct and semantically appropriate (cf. Levelt, 1989). However, recently it has been argued that such real-word phonological errors are treated in self-monitoring as phonological and not as lexical errors (Nooteboom, 2005). Also, in the current experiment syntactic and semantic criteria are useless, because there is no useful syntactic or semantic context. The finding that primed-for word–word errors can be interrupted and repaired as fast as primed for nonword–nonword errors, suggests that self-monitoring does not fully rely on a criterion of lexicality for the detection of phonological speech errors, but also has means to compare the error form with the target form, as is the case in self-monitoring of overt speech (where there is no evidence whatsoever of a criterion of lexicality; cf. Nooteboom, 2005). It seems, then, that self-monitoring of both inner and overt speech may have access to the intended word form (something denied by Levelt, 1989 and Levelt et al., 1999), and that self-monitoring of inner speech (but not of overt speech) in addition applies a quick and dirty criterion of lexicality. It should be noted here that the fast detection of lexical errors in inner speech would be no problem for production-based monitors (cf. Postma, 2000).

The assumption that self-monitoring of inner speech uses an additional criterion of lexicality makes sense if one believes that a major goal of self-monitoring of inner speech is to prevent errors from becoming public. This requires speed, and the additional criterion of lexicality may help in speeding up detection. The early interruptions studied in the present experiment are cases where the attempt to prevent a speech error from becoming public, just failed. There would be much less hurry in detecting, rejecting and repairing a speech error in overt speech. Here the goal of the action would presumably be to avoid misunderstanding being caused by the already public speech error, and the repair would contain cues for the listener that would be helpful to replace the error form with the repair form in interpretation (Levelt, 1983; Levelt, 1989).

5. Conclusions

Early overt self-interruptions and self-repairs are reactions to inner speech, not to overt speech. Self-monitoring of inner speech rejects nonwords more frequently than real words, self-monitoring of overt speech does not. The fact that phonological speech errors form real words more often than nonwords can be explained by nonwords being more often rejected from inner speech than real words. This explanation is supported by the finding that the rejection rate for nonword errors in inner speech increases with increasing phonetic distance between error and target, whereas also the overrepresentation of completed real-word errors compared to nonword errors increases with increasing phonetic distance between error and target. Self-monitoring of inner speech supposedly is speaker-oriented and has as its function to prevent errors in inner speech from becoming public. This requires speed, and therefore a quick and dirty global criterion of lexicality for error detection would be helpful. Self-monitoring of overt speech would be more listener-oriented and aims at repairing any damage to message transmission caused by already public errors.

Acknowledgement

Thanks are due to Theo Veenker for his technical support, to Eelco de Vries for running the experiment, to Hugo Quené for his help in analyzing the data, and to two anonymous reviewers for their highly valuable remarks.

References

Baars, B.J., 1980. On eliciting predictable speech errors in the laboratory. In: Fromkin, V.A. (Ed.), Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand. Academic Press, New York, pp. 307–317.

Baars, B.J., Motley, M.T. 1974. Spoonerisms: Experimental elicitation of human speech errors. Journal Supplement Abstract Service, Fall 1974. Catalog of Selected Documents in Psychology.

Baars, B.J., Motley, M.T., MacKay, D.G., 1975. Output editing for lexical status in artificially elicited slips of the tongue. J. Verbal Learn. Verbal Behav. 14, 382–391.

Blackmer, E.R., Mitton, J.L., 1991. Theories of monitoring and the timing of repairs in spontaneous speech. Cognition 39, 173–194.

Dell, G.S., 1986. A spreading-activation theory of retrieval in sentence production. Psychol. Rev. 93, 283–321.

Dell, G.S., 1990. Effects of frequency and vocabulary type on phonological speech errors. Lang. Cognitive Process. 5, 313–349.

Dell, G.S., Reich, P.A., 1980. Toward a unified model of slips of the tongue. In: Fromkin, V.A. (Ed.), Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand. Academic Press, New York, pp. 273–286.

Dell, G.S., Reich, P.A., 1981. Stages in sentence production: an analysis of speech error data. J. Verbal Learn. Verbal Behav. 20, 611–629.

Del Viso, S., Igoa, J.M., Garcia-Albea, J.E., 1991. On the autonomy of phonological encoding: evidence from slips of the tongue in Spanish. J. Psycholinguist. Res 20, 161–185.

Fromkin, V.A., 1973. The non-anomalous nature of anomalous utterances. In: Fromkin, V.A. (Ed.), Speech Errors as Linguistic Evidence. Mouton, The Hague, pp. 215–242.

Garrett, M.F., 1976. Syntactic process in sentence production. In: Wales, R.J., Walker, E.C.T. (Eds.), New Approaches to Language Mechanisms. North-Holland Publishing Company, Amsterdam, pp. 231–256.

Hartsuiker, R., Corley, M., Martensen, H., 2005a. The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: related Beply to Baars, Motley, and MacKay 1975. J. Memory Lang. 52, 58–70.

Hartsuiker, R., Kolk, H.H.J., Martensen, H., 2005b. The division of labour between internal and external speech monitoring. In: Hartsuiker, R., Bastiaanse, Y., Postma, A., Wijnen, F. (Eds.), Phonological Encoding and Monitoring in Normal and Pathological Speech. Psychology Press, Hove, pp. 187–205.

Humphreys, K., 2002. Lexical bias in speech errors. Unpublished doctoral dissertation, University of Illinois at Urbana, Champaign.

Kolk, H.H.J., 1995. A time-based approach to agrammatic production. Brain Lang. 50, 282–303.

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. The MIT Press, Cambridge, Massachusetts.

Levelt, W.J.M., Roelofs, A., Meyer, A.S., 1999. A theory of lexical access in speech production. Behav. Brain Sci. 22, 1–75.

Liss, J.M., 1998. Error-revision in the spontaneous speech of apraxic speakers. Brain Lang. 62, 342–360.

MacKay, D.G., 1973. Spoonerisms: The structure of errors in the serial order of speech. In: Fromkin, V.A. (Ed.), Speech Errors as Linguistic Evidence. Mouton, The Hague, pp. 164–194.

MacKay, D.G., 1992. Constraints on theories of inner speech. In: Reisberg, D. (Ed.), Auditory Imagery. pp. 274–285.

Motley, M.T., 1980. Verification of "Freudian" slips and semantic pre-articulatory editing via laboratory-induced spoonerisms. In: Fromkin, V.A. (Ed.), Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand. Academic Press, New York, pp. 133–148.

Motley, M.T., Baars, B.J., 1979. Effects of cognitive set upon laboratory induced verbal Freudian slips. J. Speech Hear. Res. 22, 421–432.

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. J. Verbal Learn. Verbal Behav. 21, 578–594.

Nickels, L., Howard, D., 1995. Phonological errors in aphasic naming: comprehension, monitoring and lexicality. Cortex 31, 209–237.

Nooteboom, S.G., 1969. The tongue slips into patterns. In: Sciarone, A., et al. (Eds.), Nomen, Leyden Studies in

Linguistics and Phonetics. Mouton, The Hague, pp. 114–132, Also in V.A. Fromkin, (Ed.), 1973. Speech Errors as linguistic Evidence. Mouton, The Hague, pp. 144–156.

Nooteboom, S.G., 1980. Speaking and unspeaking: detection and correction of phonological and lexical errors in spontaneous speech. In: Fromkin, V.A. (Ed.), Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand. Academic Press, New York, pp. 87–95.

Nooteboom, S.G., 2003. Self-monitoring is the main cause of lexical bias in phonological speech errors. In: Eklund, R. (Ed.), Proceedings of DiSS'03, Disfluency in Spontaneous Speech Workshop. 5–8 September 2003, Göteborg University, Sweden. Gothenburg Papers in Theoretical Linguistics 89, ISSN 0349–1021, pp. 25–28.

Nooteboom, S.G., 2005. Listening to one-self: monitoring speech production. In: Hartsuiker, R., Bastiaanse, Y., Postma, A., Wijnen, F. (Eds.), Phonological Encoding and Monitoring in Normal and Pathological Speech. Psychology Press, Hove, pp. 167–186.

Pols, L.C.W., 2004. Expanding Phonetics. In: Quené, H., Van Heuven, V. (Eds.), On Speech and Language: Studies for Sieb Nooteboom. LOT Netherlands Graduate School of Linguistics, Occasional Series, pp. 141–148.

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

Postma, A., Kolk, H., 1992. Error monitoring in people who stutter: evidence against auditory feedback defect theories. J. Speech Hear. Res. 35, 1024–1032.

Postma, A., Oomen, C.E., 2005. Critical issues in speech monitoring. In: Hartsuiker, R., Bastiaanse, Y., Postma, A., Wijnen, F. (Eds.), Phonological Encoding and Monitoring in Normal and Pathological Speech. Psychology Press, Hove, pp. 157–166.

Roelofs, A., 2005. Planning, comprehending, and self-monitoring. In: Hartsuiker, R., Bastiaanse, Y., Postma, A., Wijnen, F. (Eds.), Phonological Encoding and Monitoring in Normal and Pathological Speech. Psychology Press, Hove, pp. 42–63.

Schade, U., 1990. Kohärenz und Monitor in konnektionistischen Sprachproduktionsmodellen. In: Dorffner, G. (Ed.), Konnektionismus in Artifical Intelligence Und Kognitionsforschung. 6. Österreichische Artificial-Intelligence-Tagung (KONNAI): Proc.. Springer, Berlin, pp. 18–27.

Schade, U., 1999. Konnektionistische Sprachproduktion. Deutscher UniversitätsVerlag, Wiesbaden.

Stemberger, J.P., 1985. An interactive activation model of language production. In: Ellis, A.W. (Ed.), Progress in the Psychology of Language, Vol. 1. Erlbaum, London, pp. 143–186.