Proceedings of DiSS'03, Disfluency in Spontaneous Speech Workshop. 5–8 September 2003, Göteborg University, Sweden.
Robert Eklund (ed.), Gothenburg Papers in Theoretical Linguistics 89, ISSN 0349–1021, pp. 25–28.

Self-monitoring is the main cause of lexical bias in phonological speech errors

Sieb G. Nooteboom

UiL OTS, Utrecht University, Utrecht

Abstract

In this paper I present new evidence, stemming both from an experiment and from spontaneous speech, demonstrating that (a) lexical bias is caused by self-monitoring of inner speech, as proposed by Levelt et al. [6], and (b) that there is phoneme-to-word feedback in the mental programming of speech, as supposed by Dell [2] and Stemberger [10]. It is argued here that possibly phoneme-to-word feedback is an unavoidable side-effect of self-monitoring of inner speech.

1. Introduction

Baars, Motley & MacKay [1] elicited spoonerisms by having subjects read aloud a target like *darn bore* preceded by bias items in which at least the first phoneme in this case was a *b*, triggering the spoonerism *barn door*. They observed that the error rate for cases such as *darn bore*, triggering lexically viable outcomes, was higher than the error rate for cases like *dart board*, triggering non-word outcomes. The authors explained this result, generally known as "lexical bias", by positing an output-editing mechanism suppressing non-words from inner speech. Levelt, Roelofs & Meyer [6] supported this original explanation by Baars et al. [1] and suggested that the pre-articulatory editing leading to lexical bias is a form of covert self-correction of internal speech by the self-monitoring system that is also responsible for overt detection and correction of speech errors. A different approach has been suggested by Dell & Reich [3] and Dell [2], who proposed that lexical bias is caused by "phoneme-to-word" feedback during production processes, and therefore obviously not by the same mechanism that is responsible for the overt detection of speech errors.

The two questions I will attempt to answer in this paper are the following: (1) What is the cause of lexical bias in phonological speech errors? (2) Is there phoneme-to-word feedback in the mental programming of speech? I will do so mainly by way of an experiment adapted from Baars et al. [1], eliciting spoonerisms of the kind *darn bore* for visually presented *barn door*, or *gad boof* for visually presented *bad goof*, by phonological priming caused by preceding word pairs having the initial consonants of the spoonerism to be elicited. I made some changes to that experiment, meant to help me in answering my questions.

It occurred to me that if it would be possible to externalize some aspect of output editing, this might help us to choose between the proposed mechanisms. Now Levelt [5] (pp. 473, 474) argued that halting speech as in *v.... horizontal*, in a situation where the speaker has a choice between horizontal and vertical, cannot be a reaction to the speaker's own overt speech, because the *v* is considerably shorter than a humanly possible reaction time. So it must be a reaction to the speaker's inner speech. It is a reasonable and parsimonious assumption that this is an overt form of editing that generally stays covert, and that this is the same mechanism that is held accountable for lexical bias by Baars et al. and by Levelt et al. If so, and if we could tap such overt reactions to inner speech, it could help us decide between self-monitoring and feedback as the main mechanism causing lexical bias. The reason is that both explanations provide different predictions for the data distribution: Feedback predicts a lexical bias both in completed and in aborted spoonerisms. Output editing predicts a lexical bias in completed but an inverted lexical bias in early aborted spoonerisms. Such a result would also imply that self-monitoring of inner speech is different from self-monitoring of overt speech. Possibly one might increase the number of aborted and corrected errors in an experiment à la Baars et al., by giving the subjects more time for correction.

A further possibility to discriminate between the two explanations of lexical bias would be to assess the effect of phonetic distance between the two to-be-spoonerised consonants on the relative rates of lexical and non-lexical completed and aborted spoonerisms. The reason is that self-monitoring is supposed to rely on the same speech-comprehension system that is operative in the perception of other-produced speech. It is reasonable that single-feature errors are less easily detectable than errors involving more features. As it turns out, both feedback and self-monitoring predict an increasing lexical bias with increasing phonetic distance. But the feedback account predicts that this will be the same for completed and aborted spoonerisms. The self-monitoring account predicts that the rate of completed non-lexical errors will decrease and the rate of aborted non-lexical errors will increase with increasing phonetic distance. For this reason, potentially phonetic distance between the two to be spoonerised consonants is a helpful experimental variable.

Logically, the question what is the cause of lexical bias is separate from the question whether or not there exists phoneme-to-word feedback. So, how can we find out whether there is feedback? Feedback is supposed to generate extra activation for the units being part of the feedback loop. Extra activation will help a unit to exceed its threshold faster (cf. Roelofs [8]). Therefore feedback should affect response times in a Baars et al.-like experiment, not only response times of the errors, but, more importantly, also of the error-free productions. If the phonologically primed error is a lexical unit, the activation of phoneme nodes will be fed back to both the correct word node and the erroneous word node, and both word nodes will again re-activate their own phoneme nodes. Thus, feedback will create considerable extra activation for the phonemes of the correct word node (and also of the erroneous word node, but we assume the correct node to win out). This extra activation will shorten the response time. However, in case the primed spoonerism is non-lexical, the erroneous phoneme string has no corresponding word node, and therefore cannot help to provide extra activation for the correct string of phonemes. Therefore response times will be longer in error-free productions of word pairs primed for nonwords, than in word pairs primed for words (at least if we assume that the shortening effect of extra activation of the correct word node on the average is stronger than the delaying effect of competition between correct and erroneous word node that is also created by feedback). This reasoning has inspired a third modification of the Baars et al. experiment, enabling me to measure response times.

2. Method

The method used was basically the same as the one applied by Baars et al. [1] with some minor modifications, as explained above.

2.1. Stimuli

Priming word pairs consisted of pairs of monosyllabic Dutch words, visually presented in clear capital print on a computer screen and intended to be read silently. Before each test stimulus there were 3, 4, 5, 6 or 7 priming word pairs, chosen to prime a spoonerism, as in the sequence *give book, go back, get boot* preceding the test stimuli *bad goof*. In total there were 144 priming word pairs preceding test stimuli, and 144 non-priming word pairs preceding unprimed base-line stimuli. The initial consonants of priming word pairs and test word pairs were chosen from the set /f, s, χ, 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 3 groups of 6 stimuli with equal phonetic distance between initial consonants, viz. 1 , 2 or 3 distinctive features. For example, /f/ vs /s/ differ in 1 feature, /f/ vs. /p/ differ in 2 features, and /f/ vs. /d/ differ in 3 features. There were 36 base-line stimuli preceded by 144 non-priming word pairs and 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. After each test and each base-line stimulus word pair the subject saw on the screen a prompt SPREEK UIT (="SPEAK"). 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 7 stimuli with a variable number, on the average 4, of non-priming preceding 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 priming word pairs (144), non-priming word pairs (144 + 28 = 172), practice stimuli (7), test stimuli (36), base-line stimuli (36) and prompts (144 + 14) was 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 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 "SPREEK UIT" (="SPEAK") 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. All speech of each subject was recorded, and digitally stored on one of two tracks of DAT. On the other track of the DAT two tones of 1000 Hz and 50 ms duration were recorded with each test or filler stimulus, one starting at the onset of the visual presentation of the "SPEAK" stimulus, the other starting at the onset of the presentation of the "CORRECTION" prompt. These signals were helpful for orientation in the visual oscillographic analysis of the speech signals, and the first of these was indispensable in measuring response times.

2.4. Collecting the data

Reactions to all remaining test and filler 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, using a computer program for the visual oscillographic display and auditory playback of audio signals. Transcriptions differed in less than 2% of all utterances and in less than 10% of all utterances containing an error. Response times for all correct and incorrect responses, to both base-line and test stimuli were measured by hand in the two-channel oscillographic display from the onset of the 50 ms tone coinciding with the onset of the presentation of the visual "SPEAK" prompt to the onset of the spoken response.

3. Results

3.1. Analysis of spoonerisms

In total we found 680 erroneous reactions for primed stimuli and base-line stimuli together. Most of these errors had no relation to the experimental variables, and will not concern us here. I will concentrate on 56 completed spoonerisms, and 67 aborted spoonerisms. Do we find, as expected, a lexical bias here, and is this lexical bias the same for completed and aborted spoonerisms, or is it not? The relevant breakdown of the data is given in Table 1.

Table 1. Numbers of spoonerisms as a function of lexicality and of completed versus aborted.

	completed	aborted
lexical	37	28
nonlexical	19	39

The 56 completed spoonerisms show, as expected, a significant lexical bias (binomial test, $p<0.01$). The aborted spoonerisms, if anything, show an inverted effect of lexical bias, This is in itself not significant ($p=0.11$). However, the interaction between lexicality and completed versus aborted is significant ($\chi^2=7.21$; $df=1$; $p<0.01$). This distribution of the data rather supports a self-monitoring account of lexical bias than a feedback account. What about the effect of phonetic distance? Is there such an effect and is it the same for completed and aborted spoonerisms? The data are given in Table 2.

The main interest is in the nonlexical spoonerisms, as the self-monitoring theory predicts lexical bias from nonlexical errors being edited out more frequently than lexical ones, and also predicts that the probability of being edited out increases with phonetic distance. This is precisely what the data show. There is a strong interaction for nonlexical spoonerisms between

Proceedings of DiSS'03, Disfluency in Spontaneous Speech Workshop. 5–8 September 2003, Göteborg University, Sweden.
Robert Eklund (ed.), Gothenburg Papers in Theoretical Linguistics 89, ISSN 0349–1021, pp. 25–28.

phonetic distance and completed versus aborted, as predicted by a self-monitoring account of lexical bias.

Table 2. Numbers of spoonerisms as a function of phonetic distance in number of features between initial consonants, and of completed versus aborted, separately for lexical errors (χ^2=3.31; df=2; p>0.1; n.s) and nonlexical errors (χ^2=9.51; df=2; p<0.01; s.).

(a) lexical

	completed	aborted
1 feature	10	9
2 features	21	9
3 features	6	8

(b) nonlexical

	completed	aborted
1 feature	12	11
2 features	6	12
3 features	1	16

A priori there seems to be no reason why the data distribution would be very different for nonlexical and lexical errors, as we have no reason to assume that an effect of phonetic distance on the probability that an error is being detected in inner speech depends on lexicality. Nevertheless, the data show a very different distribution for lexical spoonerisms, with no significant interaction between phonetic distance and lexicality. This discrepancy will be taken up in the discussion.

3.2. Some additional data from spontaneous speech

An earlier study showed that neither lexical status nor phonetic distance influenced the probability of overtly correcting a spontaneous speech error (Nooteboom, [7]). The overwhelming majority of overt corrections in spontaneous speech concern posthoc corrections, where the speaker stopped after the erroneous word had been completed. Under the assumption that overt stopping during the speaking of an erroneous word is a (belated) reaction to inner speech, whereas correction after the erroneous word has been completed is a reaction to overt speech, it seems reasonable to return to the spontaneous speech data, and see whether effects of lexicality and phonetic distance can be found in speech errors where the erroneous form is interrupted. This has never been done before. Table 3 gives the relevant data for the effect of lexicality.

Table 3. Numbers of spontaneous Dutch phonological speech errors as a function of lexicality and of completed versus aborted (χ^2= 6.7; df=1; p<0.01).

	completed	aborted
lexical errors	219	18
non-lexical errors	195	35

The data show that nonlexical speech errors have a higher probability of being aborted than real-word errors, as predicted from a self-monitoring account of lexical bias. This confirms the validity of the analysis of experimentally elicited spoonerisms. These data also demonstrate that detection of errors in inner speech (aborted speech errors) differs from detection of errors in overt speech (completed speech errors), where lexicality has no effect.

3.3. Analysis of response times

Phonological priming in a Baars et al.-like experiment is supposed to create competition between correct phoneme nodes and primed phoneme nodes during the mental programming of speech. It is reasonable to expect that this competition potentially delays the firing of the winning node and thus lengthens response times in error-free productions (Cf. Roelofs [8]). This provides a way to test whether indeed response times in a Baars et al.-like experiment behave as one would expect, by comparing response times for unprimed, base-line error-free productions with response times for phonologically primed error-free productions. The average response time for base-line error-free productions is 563 ms (standard error 3.9 ms), and the average response time for primed error-free productions is 593 ms (standard error 4.4 ms). The difference is significant according to an analysis of variance with repeated measures, using a univariate design (F[1,49]=20.5; p<.001). This gives confidence in the usefulness of response times as a measure of the relative speed with which production units become available during the mental programming of speech.

In the introduction it was predicted that, assuming there is feedback between phoneme nodes and lexical nodes in the mental preparation of speech, error-free productions would have a shorter response time when the primed-for but not occurring spoonerism is lexical than when it is non-lexical. It was also predicted that this difference would increase with decreasing phonetic distance between competing phonemes. Fig. 1 gives the relevant data. An analysis of variance with repeated measures and a univariate design shows a significant main effect of lexicality (F[1,49]=18; p<.0001), a significant main effect of number of features (F[2,98]=15.5; p<.0001), and a significant interaction (F[49,1107]=12.7; p<.043). These data strongly support a model of the mental programming of speech production with feedback from phoneme nodes to lexical nodes.

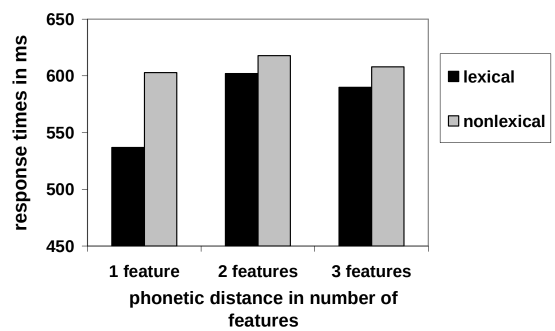

Figure 1: Response time in ms of error-free productions phonologically primed for spoonerisms, as a function of lexicality and of phonetic distance in number of features between competing phonemes.

4. Discussion

The current experiment was designed to help answering two questions: (1) What is the cause of lexical bias? and (2) Is there feedback between phoneme and word nodes in the mental preparation of speech? The data on relative frequencies of both elicited and spontaneous speech errors provided in the results section support the hypothesis by Levelt et al. [6] that lexical bias is caused by nonlexical phonological errors having a greater probability than lexical errors of being edited out from inner speech by the self-monitoring system. The data on response times support models of the mental preparation of speech exhibiting feedback from phoneme nodes to lexical nodes as proposed by Stemberger [10] and Dell [2], and as excluded by Levelt et al. [6].

As in all Baars et al.-like experiments the elicited spoonerisms are relatively few. This makes these data less convincing than one would wish. So I went looking for support from data on similar experiments. Unfortunately Baars et al. [1] and most other publications on similar experiments do not distinguish a separate category of aborted spoonerisms. Their “partial spoonerisms” apparently include such cases as *darn door* instead of *barn door*, where only the first of the two phoneme substitutions has been made. I found only one experiment, described by Humphreys [4], that is more or less comparable to the current one. She compared word-nonword with nonword-word outcomes, and found that lexical bias is completely controlled by the first word, word-nonword behaving as lexical, and nonword-word as nonlexical outcomes. Adding her numbers of lexical and nonlexical aborted spoonerisms to mine, gives 57 lexical and 77 nonlexical outcomes. This difference is as good as significant on a binomial test ($p=.0502$), providing further support for the current interpretation.

The predicted interaction between phonetic distance and completed versus aborted was only found for nonlexical errors, not for lexical errors. Conceivably this unpredicted finding is related to a different reaction of the perception system to lexical and nonlexical items. The most likely response to a nonlexical item differing only a single feature from a lexical one, is that lexical item. This probability will rapidly decrease with increasing phonetic distance. The most likely response to a lexical item differing only a single feature from another lexical item is not that other lexical item, but the item itself. This will remain the most likely response with increasing phonetic distance. This is precisely what was found.

The data on response times provide convincing evidence for the existence of phoneme-to-word feedback. This runs counter an argument by Levelt et al. [6] that there does not seem to be a function for such feedback. However, one would not need to consider such a function, if one assumes that feedback is an unavoidable side-effect of some other property of the speech production system. Levelt et al. [6] were forced by experimental evidence to introduce direct links from perception to production on three levels, lemma’s, lexemes, and phonemes. Roelofs [9] has suggested that phoneme-to-word feedback may originate from a lexeme-to-phoneme link between perception and production. This would make such a feedback an unavoidable side-effect of the way self-monitoring is organized.

5. Acknowledgement

I am grateful to my student Eelco de Vries for running the experiment, to Hugo Quené for doing the statistical analysis, and many others for helpful comments.

6. References

[1] Baars, Bernard J., Michael T. Motley & Donald MacKay. 1975. Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, vol. 14, pp. 382–391.

[2] Dell, Gary S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review*, vol. 93, pp. 283–321.

[3] Dell, Gary S. & Peter A. Reich. 1981. Stages in sentence production: an analysis of speech error data. *Journal of verbal Learning and verbal Behavior* 20, 611-629.

[4] Humphreys, Karin. 2002. *Lexical bias in speech errors*. Unpublished doctoral dissertation, University of Illinois at Urbana Champaign.

[5] Levelt, Willem.J.M. 1983. *Speaking: From Intention to Articulation.* Cambridge Massachusetts, MIT Press.

[6] Levelt, Willem. J. M., Ardi Roelofs & Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences*, vol. 22, pp. 1–75.

[7] Nooteboom, Sieb G. In press. Listening to one-self: Monitoring speech production. In: Rob J. Hartsuiker, Roelien Bastiaanse, Albert Postma & Frank.N.K. Wijnen (eds.), *Phonological encoding and monitoring in normal and pathological speech*.

[8] Roelofs, Ardi. 2002. Storage and computation in spoken word production. In: Sieb Nooteboom, Fred Weerman & Frank Wijnen (eds.), *Storage and Computation in the Language Faculty*, Dordrecht: Kluwer Academic Publishers, pp. 183–216.

[9] Roelofs, Ardi. In press. Goal-referenced selection of verbal action: modelling attentional control in the stroop task. Accepted for publication in *Psychological Review*.

[10] Stemberger, Joseph P. 1985. An interactive activation model of language production. In: Andrew W. Ellis (ed.) *Progress in the Psychology of Language* Vol 1. London: Lawrence Erlbaum (pp. 143-186).