Phonological Encoding and Monitoring in Normal and Pathological Speech

Edited by
Robert J. Hartsuiker, Roelien Bastiaanse,
Albert Postma and Frank Wijnen

Psychology Press
Taylor & Francis Group
HOVE AND NEW YORK

First published 2005
by Psychology Press
27 Church Road, Hove, East Sussex BN3 2FA

Simultaneously published in the USA and Canada
by Psychology Press
270 Madison Avenue, New York NY 10016

Psychology Press is part of the Taylor & Francis Group

Typeset in Times by RefineCatch Limited, Bungay, Suffolk
Printed and bound in Great Britain by TJ International Ltd,
Padstow, Cornwall
Cover design by Anú Design

This publication has been produced with paper manufactured to strict environmental standards and with pulp derived from sustainable forests.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
Hartsuiker, Robert J., 1968–
Phonological encoding and monitoring in normal and pathological speech / Robert J. Hartsuiker . . . [et al.].
p. cm.
Includes bibliographical references and index.
ISBN 1-84169-262-X
1. Speech disorders. 2. Linguistics. I. Title.

RC423.H345 2005
616.85′5–dc22 2004012990

ISBN 1-84169-262-X

10 Listening to oneself: Monitoring speech production

Sieb G. Nooteboom

Abstract

According to Levelt (1989) and Levelt, Roelofs, and Meyer (1999) (a) self-monitoring of speech production employs the speech comprehension system, (b) on the phonological level the speech comprehension system has no information about the lemmas and forms chosen in production, and (c) lexical bias in speech errors stems from the same perception-based monitoring that is responsible for detection and overt correction of speech errors. It is predicted from these theoretical considerations that phonological errors accidentally leading to real words should be treated by the monitor as lexical errors, because the monitor has no way of knowing that they are not. It is also predicted that self-corrections of overt speech errors are also sensitive to lexicality of the errors. These predictions are tested against a corpus of speech errors and their corrections in Dutch. It is shown that the monitor treats phonological errors leading to real words in all respects as other phonological, and not as lexical errors and that no criterion is applied of the form "is this a real word?" It is also shown that, whereas there is considerable lexical bias in spontaneous speech errors and this effect is sensitive to phonetic similarity, self-corrections of overt speech errors are not sensitive to lexical status or phonetic similarity. It is argued here that the monitor has access to the intended word forms and that lexical bias and self-corrections of overt speech errors are not caused by the same perception-based self-monitoring system. Possibly fast and hidden self-monitoring of inner speech differs from slower and overt self-monitoring of overt speech.

Introduction: Levelt's model of speech production and self-monitoring

We all make errors when we speak. When I intend to say "good beer" it may come out as "bood beer" or even as "bood gear"; or when I want to say "put the bread on the table" I may inadvertently turn it into "put the table on the table" or into "put the table on the bread". Let us call errors like "bood beer" or "bood gear", where phonemes are misplaced, phonological errors, and ones like "table on the table" or "table on the bread", where meaningful items show up in the wrong positions, lexical errors. Lexical errors supposedly arise during grammatical encoding, phonological ones during phonological encoding (Levelt, 1989). Errors as given in our examples are syntagmatic speech errors, involving two elements in the intended utterance, a source and a target, the source being the intended position of an element, the target being the position where it ends up. So in the intended utterance "bread on the table", underlying the error "table on the table", "table" is the source and "bread" the target. Speakers also make paradigmatic speech errors, involving only a single intruding element, but here I will only be concerned with syntagmatic speech errors (cf. Fromkin, 1973).

The fact that we know that speech errors exist implies that we can detect them. And we not only detect errors in the speech of others, but also in our own speech. In the collection used for the current study, roughly 50 per cent of all speech errors were detected and corrected by the speakers (an earlier analysis of Meringer's, 1908, corpus suggested somewhat higher values; Nooteboom, 1980). Apparently, part of a speaker's mind is paying attention to the speech being produced by another part of the same mind, keeping an ear out for inadvertent errors that may be in need of correction. Let us call this part of the speaking mind the "monitor", and its function "self-monitoring" (Levelt, 1983, 1989). The general question I am focussing on here is: "How is self-monitoring of speech organized, and what information does it operate on?" The question is not new. A firm stand on this issue, based on extensive empirical evidence, has been for example taken by Levelt (1989), and by Levelt et al. (1999). The reason to take their theory as a starting point is that it is the most constrained, most parsimonious, theory of speech production available. In many ways it predicts what it should and does not predict what it should not. Alternative theories will be mentioned in the discussion section.

For the present purposes the following properties of the spreading-activation theory proposed by Levelt and his associates are relevant: (1) Speech production is strictly serial and feedforward only, implying that there is no cascading activation and no immediate feedback from the level of phonological encoding to the level of grammatical encoding; (2) self-monitoring employs the speech comprehension system, also used in listening to the speech of others; (3) the speech being produced reaches the comprehension system via two different routes, the inner route feeding a covert form of not-yet-articulated speech into the speech-comprehension system, and the auditory route feeding overt speech into the ears of the speaker/listener; (4) on the phonological level there is no specific information on intended phonological forms leaking to the speech comprehension system. The monitor must make do with a general criterion of the form "is this a real word?" instead of a criterion such as "is this the word I wanted to say?"; (5) lexical bias in speech errors is caused by the same perception-based self-monitoring system that is responsible for the detection and correction of overt speech errors.

This theory leads to some predictions that can be tested by looking at properties of speech errors in spontaneous speech and their corrections. The following predictions are up for testing:

- The monitor treats phonological errors that lead to real words, such as "gear" for "beer", as lexical errors.
- If spontaneous phonological speech errors show lexical bias, as has been suggested by Dell (1986), then one should also find a lexical bias effect in self-corrections of overt speech errors.

Before testing the first prediction, it should be assessed that so-called real-word phonological errors are indeed caused during phonological and not during grammatical encoding. This question will be dealt with first. Also, it will appear below that there may be a problem in testing the first prediction, caused by the fact that many overtly corrected anticipations, such as "Yew . . . New York", may not be anticipations at all, but rather halfway-corrected transpositions. If so, there is no way of telling whether the error triggering the monitor was the real word "Yew" or the non-word "Nork" (cf. Cutler, 1982; Nooteboom, 1980). The question is whether or not this observation potentially invalidates the interpretation of a comparison between correction frequencies of phonological non-word errors, phonological real-word errors and lexical errors. It will be shown that it does. To circumvent this problem, a separate analysis will be made in which non-word and real-word phonological errors are limited to perseverations, such as "good gear" instead of "good beer", because there no part of the error can hide in inner speech. With respect to the prediction concerning lexical bias, it should be noted that reports on the existence of lexical bias in spontaneous speech errors differ. Garrett (1976) did not find evidence for lexical bias, Dell (1986) did, but Del Viso, Igoa, and Garcia-Albea (1991) did not for Spanish, although using a measure for lexical bias that is very similar to Dell's. So before studying lexical bias in self-corrections of overt speech errors, it should be assessed that there really is lexical bias in spontaneous speech errors. As will be seen, there is ample evidence for lexical bias in Dutch spontaneous speech errors. Therefore it makes sense to ask whether or not there is lexical bias in self-corrections in overt speech errors, as predicted from Levelt's theory. The reader will see that there is not. A related question is whether lexical bias is sensitive to phonetic distance between target and error phoneme, as predicted from perception-based monitoring but also from production-based theories, and if so whether the same is true for the probability of self-corrections of overt speech errors. Finally, there is the question whether the structure of the current data rather stems from a collector's bias than from the mechanisms underlying the production and perception of speech.

The following questions will now be dealt with in succession:

- Are alleged real-word phonological errors actually made during phonological or grammatical encoding?
- Does the fact that alleged corrected anticipations might sometimes have been halfway-corrected transpositions hinder the interpretation of comparisons between correction frequencies for non-word and real-word errors?
- Does the monitor treat phonological errors that lead to real words, such as "gear" for "beer", as lexical or as phonological errors?
- Do spontaneous phonological speech errors show lexical bias?
- Do self-corrections of overt speech errors show lexical bias?
- Are lexical bias and probability of self-corrections of overt speech errors equally sensitive to phonetic distance between target and error?
- Do the current data suffer from a collector's bias invalidating otherwise plausible conclusions?

Several possible explanations of the current findings will be discussed in the final section of this chapter.

The corpus

To answer the above questions two different collections of spontaneous speech errors in Dutch were used, the first collection only being used in studying lexical bias, because for these speech errors no overt self-corrections were available.

The oldest collection (AC/SN corpus) is basically the same as the one described by Nooteboom (1969). The errors were collected and noted down in orthography during several years of collecting by two people, the late Anthony Cohen and myself. Unfortunately, corrections were not systematically noted down. Collection of errors continued some time after 1969, and in its present form the collection contains some 1000 speech errors of various types, phonological syntagmatic errors outnumbering other types, such as lexical syntagmatic errors, blends, and intrusion errors. The collection was never put into a digital database and is only available in typed form, each error on a separate card. Selection of particular types of errors for the present purpose was done by hand.

The second collection (Utrecht corpus) stems from efforts of staff members of the Phonetics Department of Utrecht University, who, on the initiative of Anthony Cohen, from 1977 to 1982 orthographically noted down all speech errors heard in their environment, with their corrections, if any (cf. Schelvis, 1985). The collection contains some 2500 errors of various types, of which more than 1100 are phonological syntagmatic errors and some 185 lexical syntagmatic errors. The collection was put into a digital database, currently accessible with Microsoft Access.

Are alleged real-word phonological errors actually made during phonological or grammatical encoding?

Before making any comparisons between non-word phonological errors, real-word phonological errors and lexical errors, we have to make sure that in production alleged real-word phonological errors really arise at the level of phonological encoding and not at the level of grammatical encoding. In Table 10.1 we see confusion matrices for source and target of phonological non-word errors, phonological real-word errors and lexical errors.

These data show that in lexical errors an open-class word is never replaced by a closed-class word and a closed-class word never by an open-class word. In fact, closer analysis shows that syntactic word class is nearly always preserved (cf. Nooteboom, 1969). This is quite different for non-word phonological errors where the distribution of word-class preservation and violation is entirely predictable from relative frequencies and chance. So how do our alleged phonological real-word errors behave? Obviously they behave like non-word phonological errors, not like lexical errors. So we can be reassured that in the bulk of such errors lexical status is purely accidental. Now we are in a better position to ask whether the monitor treats real-word phonological errors as lexical errors, as predicted by Levelt et al., or rather as phonological errors. But first there is this problem with corrected anticipations perhaps being misclassified transpositions.

Corrected anticipations or halfway-corrected transpositions?

It has been observed that relatively many corrected anticipations in collections of speech errors, such as: "Yew. . . . New York", may be misclassified halfway-corrected transpositions (Cutler, 1982; Nooteboom, 1980). If we assume that speech errors can be detected in inner speech before becoming overt, in all these cases the monitor has not one but two opportunities to detect an error, and for all we know the second, hidden, part of the transposition may have been a non-word, as in the current example. This state of affairs potentially upsets any statistical differences we find in a comparison

Table 10.1 Three confusion matrices for source and target being, or belonging to, a closed- versus an open-class word, separately for phonological non-word errors, phonological real-word errors, and lexical errors

	Phonological non-word errors		*Phonological real-word errors*		*Lexical errors*	
Source	Open class	Closed class	Open class	Closed class	Open class	Closed class
Target						
Open class	303	55	169	30	135	0
Closed class	58	21	25	10	0	24

Table 10.2 Numbers of corrected and uncorrected speech errors, separately for perseverations, anticipations, and transpositions

	Perseverations	*Anticipations*	*Transpositions*	*Total*
Corrected	103	*442(?)*	*42(?)*	587
Not corrected	153	238	175	566
Total	256	*680(?)*	*217(?)*	1153

Note: There is a strong interaction between error class and correction frequency ($chi^2 = 153$; df = 2; $p < 0.001$). Cursive numbers are suspected not to correspond to what happened in inner speech. Utrecht corpus only.

between lexical and phonological real-word errors. That this is a serious threat may be shown by the following estimates of the relative numbers of anticipations and transpositions in inner speech. Let us assume that the probability of detecting an error in internal speech is not different for anticipations and perseverations (to the extent that this assumption is incorrect the following calculations will be inaccurate; but if the underlying reasoning is basically sound, they will at least provide a plausible rough estimate). We know the number of uncorrected perseverations, the total number of perseverations, and the number of uncorrected anticipations (Table 10.2). From the numbers in Table 10.2, using an equation with one unknown, one can easily calculate what the total number of anticipations, and therefore also the number of corrected anticipations, would have been, without the influx of halfway-corrected transpositions. The equation runs as follows:

103 corrected perseverations : 153 not corrected perseverations,
= ? corrected anticipations : 238 not corrected anticipations

The estimate number of corrected anticipations would then be:

$$(103 \times 238) : 153 = 160$$

The total number of anticipations would be 160 + 238 = 398. The estimate number of misclassified halfway-corrected transpositions is 442 – 160 = 282. Note that this brings the total number of transpositions in internal speech to 282 + 42 + 175 = 499 instead of 217, making transpositions by far the most frequent class of speech errors (Table 10.3). These estimates are further confirmed in the following way: The probability of remaining uncorrected is 0.6 for both perseverations and anticipations. A transposition contains an anticipation plus a perseveration. The probability of remaining uncorrected should therefore be $0.6 \times 0.6 = 0.36$. The new estimate of the fraction of transpositions remaining uncorrected equals:

$$1 - (42 + 282): 499 = 0.35$$

Table 10.3 Numbers of corrected and uncorrected speech errors in inner speech, separately for perseverations, anticipations, and transpositions

	Perseverations	*Anticipations*	*Transpositions*	*Total*
Corrected	103	*160*	*324*	587
Not corrected	153	238	175	566
Total	256	*398*	*499*	1153

Note: Cursive numbers are estimated (see text). Utrecht corpus only.

Apparently, still provided that our assumption that the probability of being detected in internal speech is the same for perseverations and anticipations was correct, both parts of the error contribute equally and independently to the probability of remaining uncorrected.

From these calculations, it is at least plausible that a great many corrected anticipations in our corpus originated as halfway-corrected transpositions in inner speech. Of course we have no way of knowing which are and which are not. In all such cases in which the error is phonological we do not know whether the error triggering the monitor was a real word or a non-word. We therefore should treat any comparison between numbers of correction for real-word and non-word anticipations with caution.

Does the monitor treat phonological errors that lead to real words as lexical or phonological?

We know that lexical errors and phonological errors are treated differently by the monitor: Both the distribution of the number of words a speaker goes on speaking before stopping to correct a speech error and the distribution of the number of words a speaker retraces in his correction is different for lexical and phonological errors (Nooteboom, 1980). Our corpus of speech errors noted down with their corrections, makes it possible to compare the distributions of the number of words spoken before the speaker stops for correction, and the number of words included in the correction, between different classes of speech errors. If Levelt et al. are right in assuming that the monitor has no way of knowing whether a particular error was made during grammatical or during phonological encoding, these distributions should be different for non-word and real-word phonological errors, and the same for real-word phonological and lexical errors.

Contrary to this prediction, Figure 10.1. suggests that the distribution of the numbers of words spoken before stopping is very similar for non-word and real-word phonological errors and rather different for real-word phonological errors and lexical errors.

To test these predictions statistically, for the moment neglecting the threat stemming from corrected anticipations being halfway-corrected transpositions, the numbers underlying Figure 10.1 were collapsed into a 2 × 2 matrix

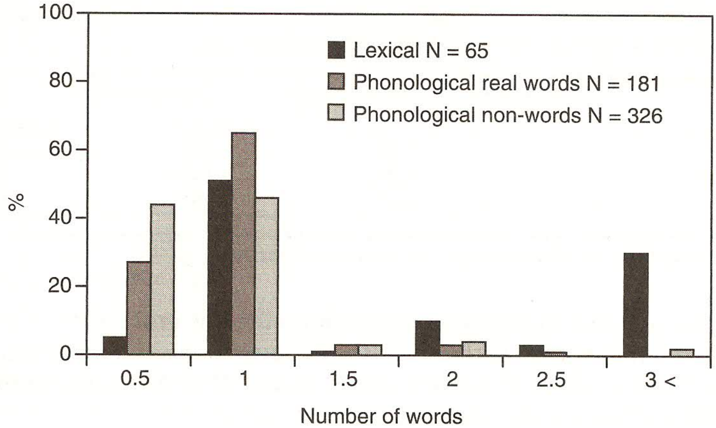

Figure 10.1 Percentage of speech errors as a function of the number of words spoken before stopping for correcting a speech error, plotted separately for lexical errors, phonological errors leading to non-words, and phonological errors accidentally leading to real words.

Table 10.4 Numbers of speech errors as a function of number of words spoken before stopping to correct a speech error, separately for non-word phonological errors, real-word phonological errors, and lexical errors

n	*1 or less*	*More than 1*
Phonological non-word	294	32
Phonological real word	163	18
Lexical	36	29

Note: Phonological non-word errors do not differ significantly from phonological real-word errors ($chi^2 = 0.00217$; df = 1; $p > 0.95$); real-word phonological errors differ significantly from lexical errors ($chi^2 = 37$; df = 1; $p < 0.0001$). Utrecht corpus only.

in order to avoid extremely small expected values, while keeping the relevant differences. The collapsed matrix is shown in Table 10.4. Phonological real-word errors differ significantly from lexical but not from phonological non-word errors. This suggests that the monitor treats the phonological real-word errors as phonological ones.

Figure 10.2 presents a similar comparison for the number of words repeated in the correction. The corresponding collapsed matrix of the underlying numbers is given as Table 10.5. Again, these data suggest that the monitor treats phonological real-word errors as phonological and not as lexical ones.

A great proportion of the data in Tables 10.4 and 10.5 concern corrected anticipations. As discussed in the previous paragraph, we should treat these data with some caution. In Tables 10.6 and 10.7 data are presented limited to phonological non-word and real-word perseverations to be compared with

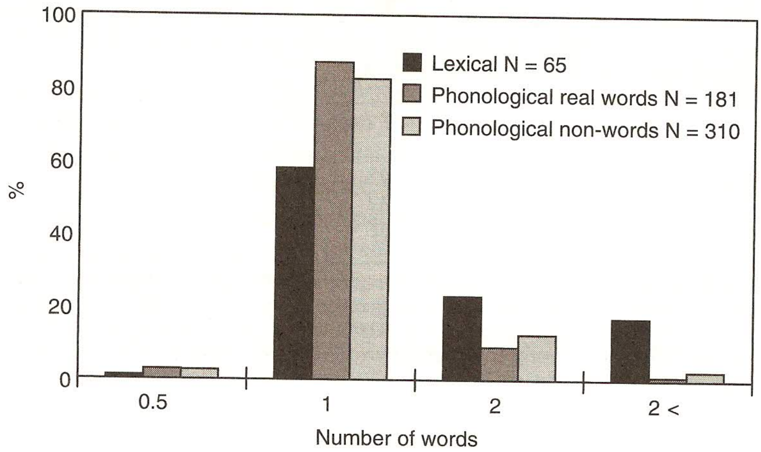

Figure 10.2 Percentage of speech errors as a function of the number of words spoken in the correction, plotted separately for lexical errors, phonological errors leading to non-words, and phonological errors accidentally leading to real words.

Table 10.5 Numbers of speech errors as a function of number of words repeated in the correction, separately for non-word phonological errors, real-word phonological errors and lexical errors

n	*1 or less*	*More than 1*
Phonological non-word	264	46
Phonological real word	161	20
Lexical	39	26

Note: Less than 1 indicates that the speaker not even went back to the beginning of the word containing the error. This occurred only in compounds. Phonological non-word errors do not differ significantly from phonological real-word errors ($chi^2 = 1.41$; df = 1; $p > 0.1$); real-word phonological errors differ significantly from lexical errors ($chi^2 = 26$; df = 1; $p < 0.0001$). Utrecht corpus only.

Table 10.6 Numbers of speech errors as a function of number of words spoken before stopping for correction, separately for non-word phonological perseverations, real-word phonological perseverations, and lexical errors

n	*Less than 1*	*1*	*More than 1*
Phonological non-word	24	25	0
Phonological real word	15	11	0
Lexical	4	32	29

Note: Less than 1 indicates that the speaker did not complete the word containing the error. Phonological non-word perseverations do not differ significantly from phonological real-word perseverations ($chi^2 = 0.52$; df = 1; $p > 0.3$); real-word phonological perseverations differ significantly from lexical errors ($chi^2 = 35$; df = 2; $p < 0.0001$). Utrecht corpus only.