Journal of Memory and Language 68 (2013) 26–38

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Heft lemisphere: Exchanges predominate in segmental speech errors

Sieb G. Nooteboom *, Hugo Quené

Utrecht Institute of Linguistics OTS, Netherlands

ARTICLE INFO

Article history:
Received 11 November 2011
revision received 24 August 2012
Available online 30 September 2012

Keywords:
Speech errors
Speech planning
Self-monitoring
Inner speech
Scan-copier

ABSTRACT

In most collections of segmental speech errors, exchanges are less frequent than anticipations and perseverations. However, it has been suggested that in inner speech exchanges might be more frequent than either anticipations or perseverations, because many half-way repaired errors (Yew…uhh..New York) are classified as repaired anticipations, but may equally well be half-way repaired exchanges. In this paper it is demonstrated for experimentally elicited speech errors that indeed in inner speech exchanges are more frequent than anticipations and perseverations. The predominance of exchanges can be explained by assuming a mechanism of planning and serial ordering segments during the generation of speech that is qualitatively similar to the scan-copier model proposed by Shattuck-Hufnagel (Sublexical units and suprasegmental structure in speech production planning. In P.F. MacNeilage (Ed.), *The production of speech (pp. 109–136)*. New York: Springer).

Introduction

Errors of speech come in many varieties. One major distinction is between syntagmatic errors and paradigmatic errors. In syntagmatic errors there is a source and a target within the utterance, as in *heft..left hemisphere* (most English examples in this paper are taken from Fromkin,1993), where supposedly the source of the error is the *h* of *hemisphere*, the target is the position of the *l* of *left*, and an intruding segment, taken from the source, is misplaced into the target position. In paradigmatic errors the source of the error is to be sought outside the utterance, as in *on your left..uh your right hand*. In this paper the focus will be on syntagmatic speech errors. Another major distinction is between errors where the misplaced units are meaningful lexical units, such as in the last example, or meaningless speech sounds as in *some kunny kind* instead of *some funny kind*. In collections of speech errors syntagmatic segmental errors far outnumber syntagmatic lexical errors, by a factor of 5 or 6 (e.g. Nooteboom, 1973; Nooteboom, 2005a). Here syntagmatic segmental errors take center stage. The misplaced unit in a speech error cannot only replace another unit as in the examples given, but also be omitted or added, as in *acon and begs* for *bacon and eggs*, where the *b* is omitted from *bacon* and added to *eggs*.

Another dimension that is relevant to the study of syntagmatic segmental speech errors, is what may be called the "direction" of the error. A speech sound may come too early, as in *a Tanadian from Toronto* for *a Canadian from Toronto*. Such errors we call "anticipations". Or a speech sound may come too late, as in *she can she it* instead of *she can see it*. These errors are named "perseverations". And a speech error may exchange two speech sounds as in *teep a cape* for *keep a tape*. These are sometimes called "transpositions". Here they will simply be called "exchanges". In collections of speech errors in spontaneous speech exchanges are most often less frequent than anticipations and perseverations. Typically, Nooteboom (1973) found in a corpus of errors in spontaneous Dutch collected by Cohen (1966), 78% anticipations 15% perseverations, and only 7% exchanges. Nooteboom (1980), counting errors in Meringer's (1908) corpus, found 61% anticipations, 28% perseverations, and 11% exchanges. Nooteboom (2005a) reporting on the much larger Utrecht corpus of

* Corresponding author. Address: Cor Ruyslaan 20, 3584 GD Utrecht, Netherlands.

E-mail address: S.G.Nooteboom@uu.nl (S.G. Nooteboom).

0749-596X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jml.2012.08.004

speech errors (Schelvis, 1985), mentions 60% anticipations, 22% perseverations and 18% exchanges. Although these numbers seem to suggest that exchanges are less frequently made than anticipations and perseverations, this may not be true for the generation of sound errors in inner speech. The reason is that in all the studies mentioned, errors of the form *heft*… *left hemisphere* are classified as repaired anticipations. Both Nooteboom (1980, 2005a) and Shattuck-Hufnagel (1979, 1983) pointed out that such incomplete errors, corrected in midstream, may be incipient anticipations, as *heft hemisphere*, but may also be the first parts of exchanges, as *heft lemisphere*. Shattuck-Hufnagel (1979) suggested that there "is evidence that exchange errors are more common than substitutions" (p. 323), referring to an analysis of the MIT-CU corpus that seems to suggest that feature constraints on exchanges and substitutions are quite different, and that incomplete errors are indistinguishable from exchanges but significantly different from substitutions in their feature constraints (p. 325; also see Shattuck-Hufnagel & Klatt, 1979). Nooteboom (2005a) concluded, on the basis of a somewhat speculative argumentation, that in inner speech exchanges are probably more frequent than both anticipations and perseverations.

A possible argumentation for the latter claim runs as follows. Imagine that an anticipatory error, like *heft hemisphere* is made in inner speech. In that case the monitor watching out for speech errors in inner speech has only one single chance to detect the error, by detecting the erroneous form heft. However, when an exchange error, such as *heft lemisphere*, is being made in inner speech, there are two erroneous forms that can trigger error detection, *heft* and *lemishere*. One may note, of course, that this presupposes that an error can be detected in inner speech, before this error is spoken. As it happens, this appears not only to be true for hidden errors such as *lemisphere* in *heft..left hemisphere*, but also for most, if not all, of the overt incomplete errors. One reason is that these incomplete errors in a great many cases are fragments of speech consisting of only a single word-initial consonant, or a word-initial CV-combination, as in *d..barn door* or *ga..bad goof*. Such fragments of speech generally are shorter than a humanly possible reaction time (Nooteboom, 2005b). Therefore error detection leading to the command to stop speech, must have taken place before the error was made overt. A second reason is that the interval between offset of the interrupted, incomplete error and onset of the repair very often is in the order of 0 ms, suggesting that not only error detection but also error repair was planned before the spoken realization of the error (Blackmer & Mitton, 1991; Nooteboom, 2005b). It has also been shown experimentally that monitoring inner speech for speech errors by the speaker is faster and more efficient than monitoring overt speech (Hartsuiker, Kolk, & Martensen, 2005).

Assuming, then, that indeed exchanges have two chances to be detected against anticipations only one, it is reasonable to conclude that most incomplete, interrupted, errors stem from exchanges, and only a minority from anticipatory errors. This has been argued by Nooteboom (2005a) for errors in spontaneous speech. The weakness of Nooteboom's claim was that in his collection of segmental speech errors in spontaneous Dutch most exchanges supposedly having occurred in inner speech remained hidden in the large set of early interrupted errors of the type *Yew*…*uhh..New York*. They could not be distinguished from early repaired anticipations, and could therefore not actually be counted. In this paper we will describe an experiment set up with the explicit purpose of severely reducing the number of early interrupted speech errors, and of explicitly eliciting not only exchanges but also anticipations and perseverations. If our idea that in inner speech exchanges are considerably more frequent than anticipations and perseverations is correct, than we will find that exchanges predominate not only when these are explicitly elicited but also when anticipations and perseverations are explicitly elicited. This is strongly suggested by results obtained by Karen Humphreys and described in her unpublished dissertation (2002). She did an experiment explicitly eliciting anticipations and perseverations instead of exchanges and found that, when anticipations were explicitly elicited, nevertheless unrepaired exchanges were more frequently made than unrepaired anticipations. Results were less clear when perseverations were explicitly elicited, probably because of the scarcity of segmental speech errors in that experiment. However, the most frequent type of segmental speech error was what she called "aborted onset exchange", and what we call "early interruptions", that derive either from exchanges or from anticipations in inner speech. This means that in her data there remains some uncertainty whether indeed exchanges had been the most frequent type of segmental errors in inner speech. In another experiment Humphreys compared numbers of segmental errors when exchanges were explicitly primed with those obtained when anticipations were explicitly primed. In this experiment priming exchanges was far more effective than priming anticipations in eliciting segmental speech errors, suggesting that, at least in experiments explicitly priming segmental speech errors, exchanges are not caused solely by an initial anticipation automatically followed by an anticipation (as suggested by Shattuck-Hufnagel, 1979, 1983). Here again unrepaired exchanges were more frequent than unrepaired anticipations. But again, as in collections of segmental speech errors in spontaneous speech, early interruptions were more frequent than any other type of segmental speech error. This made the ratio between exchanges and anticipations in inner speech invisible.

As it happens, the predominance of exchanges in inner speech is not, at least not quantitatively correctly, predicted from existing models of serial ordering of segments in speech production. Dell's computational spreading activation model (1986) has a feature promoting exchanges, viz. post-selection inhibition of activation, but the resulting effect is too weak to predict a strong predominance of exchanges. Parallel Distributed Models of speech production, as exemplified by the model proposed by Dell, Juliano, and Govindjee (1993), do not generate segmental exchanges at all. The computational spreading activation model WEAVER++, as described by Levelt, Roelofs, and Meyer (1999), does not have a mechanism that would more or less automatically generate a following perseveration after an anticipation has occurred, and would thus not easily generate exchanges. Dell, Ferreira, and Bock (1999), in their commentary on the 1999 article by Levelt et al. describe a possible mechanism to explain errors such as *sed rock* for *red sock*. According to Dell et al. "(f)irst, a hyperactivated unit [s] is selected over the correct one, [r]. The selected unit is then inhibited and hence is less likely to be selected for the next syllable. But the replaced unit [r] is still available for that syllable because it was not inhibited". This view of a mental process potentially generating exchanges is basically similar to the "scan-copier model of speech production" proposed by Shattuck-Hufnagel (1979, 1983).

Shattuck-Hufnagel assumes that an exchange is caused by only a single error in the process of serial ordering, viz. an error of selecting the wrong unit because it is accidentally more activated than the correct one. Because this wrongly selected unit is "checked off" as being used (i.e. it is de-activated or inhibited) the correct unit, still not being used and therefore not checked-off, is automatically inserted in the position that was reserved for the unit that now is inadvertently not available any more. Within her view, whereas an exchange results from only a single slip in the process of serial ordering speech segments, an anticipation results from two such slips, first a selection error and then an error of not de-activating the selected unit. A perseveration also would result from two consecutive errors, viz. an error of not de-activating a correctly inserted segment followed by inserting that segment for a second time in a fitting position, replacing a correct segment. When we assume that the two consecutive errors are independent, Shattuck-Hufnagel's view of the process of serial ordering predicts an abundance of exchange errors (Shattuck-Hufnagel herself has refrained from drawing this conclusion from her model). However, as has been pointed out to us by Gary Dell (personal communication), this would make segmental anticipations and perseverations very rare indeed, compared to segmental exchanges. If each of the two types of errors proposed by Shattuck-Hufnagel would have a frequency of 1 in a thousand words, which is not unrealistic, exchanges would occur with this frequency, but both anticipations and perseverations would only have a frequency of 1 in a million words. This seems hardly realistic. We conclude that, if our assumption that exchanges predominate in inner speech is correct, current models do not make realistic predictions as to the relative frequencies of exchanges, anticipations and perseverations. We will come back to this in the discussion.

The models mentioned above imply the classical assumption that segmental speech errors consist of omitting, adding, or substituting complete speech segments or phonemes, on the level of speech planning, supposedly coinciding with "internal" or "inner" speech (cf. Levelt, 1989; Levelt et al., 1999). However, more recently it has been found by Goldstein, Pouplier, Chen, Saltzman, and Byrd (2007) and also by McMillan and Corley (2010) that so-called segmental errors of speech, at least under the conditions of the elicitation experiments reported, on the level of articulation more often than not show simultaneous articulatory gestures stemming from apparently simultaneously activated competing gestural units. Of course this is an important result for our ideas on the process of serial ordering in speech production. Pouplier and Goldstein (2005) also demonstrated that our perception of segmental speech errors and therefore also our transcription of those errors, often is erroneous, because simultaneous and conflicting articulatory gestures are often heard as either single correct or as single erroneous segments. According to McMillan and Corley (2010) the implication is that each categorization in terms of 'correct' and 'error', as is done in transcribing speech errors, may be artificial. Description should rather be in terms of the variation in articulatory movement caused by interference between two segments that simultaneously affect articulation. In this paper we do not follow this advice. Here we define an overt speech error as an auditorily perceivable mispronunciation. We admit that we do not know the articulatory nature of such speech errors, in particular whether these speech errors are gradient or not, and whether they in articulation demonstrate the activation of simultaneous units competing for the same slot, but we prefer to limit speech errors to those cases where in principle listeners can observe a speech error. In this way we can also relate our data to self-monitoring of inner speech, making a speaker perceive and detect his or her own speech errors even before these are made. This is important because, as the reader may remember, the scarcity of exchanges in collections of speech errors in spontaneous speech we explain from the filtering function of self-monitoring inner speech. The experiment to be described below was set up such that the quantitative effects of self-monitoring were artificially reduced, so that the supposed predominance of exchanges in inner speech would more clearly also cause a predominance of exchanges in the overt segmental errors.

The supposed predominance of segmental exchanges in inner speech does in most situations not correspond with a predominance of exchanges in overt speech errors. We submit that this is so because segmental exchanges are more often detected and repaired than segmental anticipations and perseverations (cf. Nooteboom, 2005a). This would be an effect of self-monitoring: Exchanges presumably stand a higher chance of being filtered out by self-monitoring before speech starts than anticipations and perseverations. Therefore the number of overt exchanges underestimates the number of exchanges made in inner speech before self-monitoring has been applied. If this reasoning is valid, one also expects in an experiment eliciting segmental speech errors a difference between exchanges on the one hand and anticipations and perseverations on the other in response times. The reason is as follows: It has been shown that, in an experiment eliciting segmental speech errors, utterances containing a speech error have longer response times than correct utterances. This increase in response times for speech errors can be ascribed to self-monitoring (Nooteboom & Quené, 2008): Each segmental speech error is a suspicious, attention requiring item and thus costs extra time, whether or not the error is detected and repaired. Because supposedly self-monitoring focuses on inner, not on overt speech, this extra time is spent before the error-containing utterance is spoken, and this would be reflected in the response time defined as the time period between a cue to pronounce a particular stimulus word pair and the onset of the corresponding utterance. If an utterance does not only contain the elicited speech error, but also an additional speech error, as for example in *barn door* turning not into *darn bore*, but into *dark boat* this would cost extra time. This was demonstrated to be the case by Nooteboom and Quené (2008). We predict that this effect of an additional error will also appear in the experiment to be described below, supporting our view of self-monitoring. This is relevant for the current purpose: A segmental exchange consists of two consecutive segmental speech errors and each anticipation and perseveration consists of only a single segmental speech error. As each segmental speech error is a suspicious item that requires attention by the monitor, and therefore some time, we predict that response times for segmental exchanges are longer than those for segmental anticipations and perseverations. If this prediction is confirmed, this would support the current view that the actual distribution of the three types of segmental speech errors in overt speech is to a large extent controlled by self-monitoring inner speech.

It should be noted that if Goldstein et al. (2007) and McMillan and Corley (2010) are right, then in inner speech underlying the actual articulatory gestures they measured, in the case of a segmental speech error a single slot for a speech segment can be filled by two simultaneous competing segments, the degree of relative activation of these two competing segments potentially varying between dominance of the erroneous segment and dominance of the correct segment, with in between a region where there is ambiguity between the two segments. If their results can be generalized to other situations, then this would also be the case in other experiments eliciting segmental speech errors including the current experiment. If so, then with respect to positions primed for segmental errors, self-monitoring inner speech for speech errors by the participants is confronted with (a) unambiguous correct segments, (b) segments that are ambiguous between correct and erroneous segments, and (c) unambiguous erroneous segments. Those cases where a segmental speech error is heard and transcribed by the experimenter potentially contain both unambiguous and ambiguous erroneous segments. It is reasonable to assume that unambiguous error segments are more often detected in self-monitoring than ambiguous error segments. This cannot be shown directly of course, because we do not have independent access to the degree of ambiguity of segmental information in inner speech. But we do know that perceptual ambiguity leads to response conflict and that response conflict leads to an increase in reaction time (Botvinick, Braver, Barch, Carter, & Cohen, 2001; see also Szmalec et al., 2008). If the two assumptions (1) that unambiguous error segments are more often detected than ambiguous ones and (2) that ambiguity leads to increased reaction time are both correct, then we can predict that those speech errors that are detected and repaired by the participants, being less ambiguous, on average have shorter reaction or response times than those speech errors that are not detected and repaired by the participants, because the latter will on averaged be more ambiguous. If such a relation between detection and response times is actually found, this would support the idea that segmental positions in inner speech can be ambiguous between a correct and an erroneous segment, as suggested by the results obtained by Goldstein et al. (2007) and McMillan and Corley (2010).

An experiment

There are three main reasons to revert to an experimental approach in studying the alleged predominance of exchanges as compared to anticipations and perseverations in the planning of serial ordering of speech segments. The first reason is that past experience shows that the overrepresentation of interrupted speech errors relative to exchange errors in standard elicitation experiments (cf. Baars & Motley, 1974) is less than found in segmental errors in spontaneous speech. In the experiments reported by Nooteboom and Quené (2008) numbers of completed spoonerisms and interruptions are in the same range. This holds the promise that in an appropriate experiment, in which the probability of early interrupted errors is further reduced, the underlying distribution of error types will shine through more clearly than in collections of errors made in spontaneous speech.

The second reason is as follows. If indeed, as suggested here, the process of planning and serial ordering of speech segments is such that exchanges are more easily made than anticipations and perseverations, then one would expect (1) that in explicitly eliciting exchanges very few anticipations and perseverations will be made, and (2) that in explicitly eliciting anticipations or perseverations, relatively many exchanges will be made. The first expectation is already borne out by data obtained in earlier experiments. For example, Nooteboom and Quené (2008), describing two experiments eliciting segmental exchanges, in their Experiment 1 found only 6 anticipations against 86 completed exchanges and 108 interruptions, and in their Experiment 2 only 1 anticipation against 54 completed spoonerisms and 49 interruptions. Clearly, when exchanges are explicitly elicited, anticipations are rare events. The other expectation, viz. that exchanges are relatively frequent when anticipations or perseverations are explicitly elicited, is supported by data obtained by Humphreys (2002), under the assumption that most of the relatively frequent early interruptions in her experiments derive from exchanges in inner speech. The main purpose of the experiment reported here is to demonstrate the predominance of segmental exchanges when the frequency of early interruptions is reduced.

The third reason for setting up an experiment is that this makes it possible not only to use relative frequencies of error types as dependent variable, but also to measure response times. As explained in the introduction, we have made three predictions on response times: (1) Utterances containing an elicited error plus an additional speech error have longer response times than utterances containing only the elicited speech error. (2) Exchanges have longer response times than anticipations and perseverations. (3) Utterances containing segmental speech errors that are detected and repaired have shorter response times than utterances containing undetected speech errors. These predictions will be tested below.

Method

The method of this experiment was taken from the so-called SLIP (Spoonerisms of Laboratory-Induced Predisposition) technique introduced by Baars and Motley (1974). Basically this technique consists of presenting participants visually with one word pair at the time. Each word pair is to be read silently until a particular word pair, the target or stimulus word pair, is followed by a cue, for example a string of question marks, signaling that the last word pair seen is to be spoken aloud. The precursor word pairs are chosen such that they phonologically prime the stimulus word pair for a particular segmental speech error, mostly an exchange of initial consonants. The method in the current experiment was derived from the one used in Experiment 1 of Nooteboom and Quené (2008), with a number of modifications necessary to elicit not only exchanges but also anticipations and perseverations in a controlled way (see Table 1 below). One other major difference between the experiment by Nooteboom and Quené (2008) and the current experiment was the following. Nooteboom and Quené, next to a test condition in which stimuli were phonologically primed for exchanges by sets of precursors, also used a base-line condition in which the same stimuli were preceded by sets of non-priming precursors. In the current experiment the base-line condition was omitted. This was done because the previous experiments showed that in the base-line condition the number of speech errors of the same type as primed for in the test condition, was negligible, so the base-line can appropriately be set at zero errors. Other modifications are related to the goal of eliciting as many as possible segmental errors of the three different types, anticipations, perseverations and exchanges, and reducing the frequency of early interrupted errors. The method is described in detail below.

Stimulus material

There were two stimulus lists, each with 12 stimulus word pairs primed for exchanges, 12 word pairs primed for anticipations, 12 word pairs primed for perseverations, and 46 filler word pairs preceded by a varying number of non-priming precursors. For the test stimulus word pairs priming was obtained by having each stimulus word pair preceded by 5 precursors, as exemplified in Table 1. Stimulus word pairs primed for perseverations were derived from ones priming for anticipations by exchanging word 1 and word 2 together with their precursors. Stimuli primed for exchanges were made phonologically similar to those primed for anticipations and perseverations. The 46 filler stimuli and the 12 test stimuli primed for exchanges occurred in both stimulus lists. The 12 test stimuli primed for anticipations in one stimulus list were turned into 12 test stimuli primed for perseverations in the other stimulus list, and vice versa. The two lists together contained 24 test stimuli primed for exchanges, 24 test stimuli primed for anticipations, 24 test stimuli primed for perseverations, and 92 filler stimuli.

It has been found in earlier similar experiments that a number of stimulus properties decrease the probability that an error is interrupted, and thereby increase the probability of completed primed-for speech errors. This, of course, would be favorable for the current experiment. It was therefore attempted to achieve these goals by (1) making the vowels in both words of each stimulus word pair and each immediately preceding priming word pair identical (cf. Dell, 1986; Nooteboom & Quené, 2008); (2) making the initial consonants phonetically similar; these always differed in only a single distinctive feature, being either voiced vs voiceless, or bilabial vs alveolar vs dorsal, or plosive vs fricative (Fromkin, 1993; Nooteboom, 1973; Nooteboom, 2005b; Nooteboom & Quené, 2008 and others); (3) avoiding stimulus word pairs where the outcome of the elicited speech error would be nonlexical (Baars & Motley, 1974; Dell, 1986; Hartsuiker, Corley, & Martensen, 2005; Nooteboom & Quené, 2008, and others); (4) exerting time pressure on the participants (Nooteboom & Quené, 2008; see procedure below); (5) making it difficult for the participants to anticipate the word pair to be spoken aloud (Nooteboom & Quené, 2008) by employing a great number of filler stimuli, each of these being preceded by 0, 1, 2, 3 or 4 non-priming word pairs. There were in each stimulus list 18 filler stimuli with 0, 7 filler stimuli with 1, 13 filler stimuli with 2, 4 filler stimuli with 3, and 4 filler stimuli with 4 precursors. The test stimulus word pairs used in the experiment are given the Appendix A.

Procedure

Each participant was tested individually in a sound treated booth. The timing of visual presentation on a computer screen was computer controlled. The order in which test and filler stimuli, along with their priming or non-priming preceding word pairs, were presented was randomized and different for each pair of an odd-numbered and the following even-numbered participant. The order of the stimuli for each even-numbered participant thus was basically the same as the one for the immediately preceding odd-numbered participant, except that

Table 1
Examples of 3 stimulus word pairs, each preceded by 5 precursors, priming for an exchange, an anticipation or a perseveration.

Precursor nr.	Exchange		Anticipation		Perseveration	
	Word 1	Word 2	Word 1	Word 2	Word 1	Word 2
1	kor	pit	peet	doof	doof	peet
2	kijf	puit	pel	nis	nis	pel
3	koet	pop	poog	wed	wed	poog
4	kuur	poel	pep	gil	gil	pep
5	kar	pak	pang	rat	rat	pang
Stimulus	paf	kap	kak	pal	pal	kap

anticipation-outcome stimuli and perseveration-outcome stimuli were interchanged. Forty-eight participants were, after the practice word pairs, presented with List 1 immediately followed by List 2, 48 other participants were presented with List 2 immediately followed by List 1. After the final word pair of each trial a "?????"-prompt, meant to elicit pronunciation of the last word pair seen (the target or stimulus word pair), was visible during 900 ms and then immediately followed by a simultaneous loud buzz sound and blank screen, both of 100 ms duration. The participants were strongly encouraged to speak the last word pair seen before this buzz sound started. This was practiced during the practice items. The buzz sound was immediately followed by a cue consisting of the Dutch word for "correction", visible during 900 ms, again followed by 100 ms with a blank screen. The participants were instructed to correct themselves immediately whenever they made an error. It was not necessary to wait for the "correction"- prompt. After the correction period and a 100 ms resetting period, the first word pair of the following trial sequence was presented. All speech of each participant was recorded with a Sennheiser ME 50 microphone, and digitally stored on disk with a sampling frequency of 48,000 Hz. The resulting speech was virtually always loud and clear. On a separate track a tone of 1000 Hz and 50 ms duration was recorded with each target stimulus word pair, starting at the onset of the visual presentation of the "?????"-prompt. These tones were helpful for orientation in the visual oscillographic analysis of the speech signals and for measuring response times. Whereas Baars, Motley, and MacKay (1975) had their participants listen to white noise during the experiment, probably to make them focus on inner speech rather than overt speech, this was avoided in the current experiment. Testing took approximately 16 min for each participant.

Scoring the data

Responses to all test and stimulus presentations were transcribed either in orthography, or, where necessary, in phonetic transcription by the first author using a computer program for the visual oscillographic display and auditory playback of audio signals. Responses were categorized as:

1. Fluent and correct responses of the type *barn door* > *barn door* or *bad goof* > *bad goof*.
2. Completed Exchanges of the type *barn door* > *darn bore* or *bad goof* > *gad boof*, with or without any additional error.
3. Completed Anticipations of the type *barn door* > *darn door*, with or without any further error.
4. Completed Perseverations of the type *barn door* > *barn bore*, with or without any further error.
5. Interrupted exchanges or anticipations.
6. Other speech errors including hesitations.
7. No response.

There were very few interruptions after the first vowel of the elicited spoonerisms (cf. Nooteboom, 2005b). All interruptions were included. Response times for all correct and incorrect responses, to both test stimuli and filler stimuli, were measured by hand in a two-channel oscillographic display from the onset of the visual prompt (= the onset of the 50 ms tone) to the onset of the spoken response. The onset of the spoken response was in most cases defined as the first visible increase in energy that could be attributed to the spoken response. However, the voice lead in responses beginning with a voiced stop was ignored because in Dutch duration of the voice lead appears to be highly variable and unsystematic both between and within participants (Van Alphen, 2004), as confirmed by a range from 0 to roughly 130 ms observed for voice leads in the current experiment.

Results

Evidence from frequencies of error types

The frequencies of observed speech errors in the earlier mentioned categories are given in Table 2 below. It is immediately clear from Table 2 that our attempt to reduce the frequency of interrupted errors was successful, and that exchanges, no longer in great numbers hidden in the class of interrupted errors, predominate over anticipations and perseverations. The frequencies of the various types of segmental speech errors were analyzed by means of bootstrapped multinomial logistic regression, similar to the analyses reported by Nooteboom and Quené (2008). The error rate in each response category was analyzed by means of multinomial logistic regression which takes into account the interdependency of responses over categories. The random effects of participants and items were simulated by means of two-stage bootstrap replications of the multinomial regression (Efron & Tibshirami, 1993; Shao & Tu, 1995, p. 247 ff). In the first stage, a sample of 59 items was drawn out of the 60 items in this experiment. (In this stage we sampled items and not participants, because the between-item variance was considerably larger than the between-participant variance, cf. Nooteboom & Quené, 2008). In the second stage, a bootstrap sample was drawn from the responses to those items selected in the first stage. No-response cases were excluded. The resulting data set was then analyzed by means of multinomial logistic regression with the priming condition (primed for exchange, for anticipation, or for perseveration) as fixed predictor, and the above-mentioned response categories as

Table 2
Frequencies of response categories obtained in the experiment, broken down by priming condition ($n = 2304$ per row).

Priming condition	Response category						
	exch	antic	persev	interr	other	no resp	correct
Exch	142	26	3	37	119	14	1963
Antic	65	43	4	25	130	24	2013
Persev	21	11	17	12	161	16	2066

the dependent variable, with correct fluent responses as the reference category. Regression was done with the function multinom in the package nnet for R (R Development Core Team, 2011; Venables & Ripley, 2002). This bootstrap-and-regression procedure was repeated 250 times. The resulting coefficients may be regarded as estimated cell means (in log odds), based on varying items and participants for each replication. Differences between these cells were evaluated by means of sign tests of the estimated means, using Bonferroni adjustment for multiple comparisons (Nooteboom & Quené, 2008).

As Fig. 1 shows, priming for exchanges is much more successful than priming for anticipations (sign test, $p < .0001$) or for perseverations ($p < .0001$). In other words, priming for exchanges yields a far larger number of exchanges than the numbers of anticipations and of perseverations in their respective priming conditions. This confirms a result obtained by Humphreys (2002), who also found that priming for exchanges is more effective than priming for anticipations. Because in priming exchanges two positions are primed for a segmental speech error, whereas in priming anticipations or perseverations only a single position is primed for a speech error, it would be reasonable to expect priming exchanges to be twice as strong as priming anticipations or perseverations. From this argument one may predict that, other things being equal, the numbers of exchanges should be twice the number of anticipations and perseverations, the latter two being equal. Note that the observed distribution of all errors in these response categories (228:80:24) suggests a 9:3:1 distribution that deviates considerably from the expected 2:1:1 distribution; likewise, the successfully primed errors (142:43:17) also suggest a similar 9:3:1 distribution. Obviously, the overrepresentation of exchanges

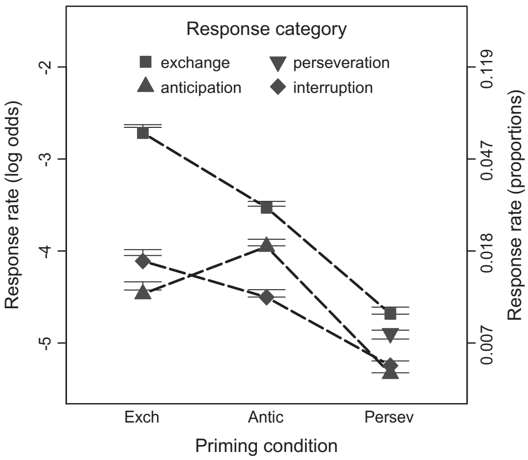

Fig. 1. Observed and estimated response rates in log-odds units, broken down by primed-for response category. Error bars correspond to 95% confidence intervals of the bootstrapped logistic regression coefficients, over 250 replications. Note that the observed response rates may deviate from the center of the bootstrap confidence intervals. The data points for produced perseverations in the priming conditions for exchanges and anticipations have been omitted because there were very few responses. The response categories "other errors" and "no responses" have also been omitted.

is not only due to the effects of their priming being stronger, but it also suggests that the underlying production processes favor exchanges over other segmental errors.

Secondly, Fig. 1 shows that exchange errors are the most frequent errors, more frequent than anticipations or perseverations, even when the priming condition should in fact elicit anticipations (triangles; sign test, $p < .0001$) or perseverations (inverted triangles; $p < .0001$). Nevertheless, the response rates summarized in Fig. 1 clearly show that priming does yield a positive effect on the number of elicited speech errors in the primed-for category. Finally, the error rate of successfully elicited anticipations (43) is significantly higher than that for successfully elicited perseverations (17), as confirmed by the bootstrap results in Fig. 1 (sign test, $p < .0001$). This is in line with reports on error frequencies in spontaneous speech (Cohen, 1966; Nooteboom, 1973; Rossi & Peter-Defare, 1998, and others) and also with predictions from the theory of speech planning and production (Levelt et al., 1999) and with predictions for normal and slow speech from Dell's spreading–activation model (Dell, 1986).

Evidence from response times

With respect to response times, we have made three predictions. One is that response times will be longer when segmental errors are accompanied by an additional error, because rejection of the elicited error and replacing this with another error, takes time. A second is that exchanges have longer response times than anticipations and perseverations. This would be so because the monitor for speech errors encounters not one but two suspicious, attention requiring, items. Note that it is assumed here that such erroneous segments increase response times, because they require some attention from the monitor, also when they are not detected and repaired. A third prediction was that unrepaired errors have longer response times than repaired errors, because on average unrepaired errors in inner speech will probably be more ambiguous than repaired errors, therefore causing more response conflict.

In order to verify these predictions, response times were log-transformed and then analyzed by means of mixed-effects regression analyses. In the first analysis, two fixed factors were included in the regression model, using dummy factors: first, the response category (excluding "fluent" and "other error" responses, which yielded for the present purpose irrelevant response times). The classification of responses with vs. without an additional error yielded the second fixed factor. Unfortunately, there were so few repaired errors (21 repaired exchanges, 10 repaired anticipations, 3 repaired perseverations) that these were excluded from the first analysis; the contrast between response times for repaired and unrepaired responses was investigated in a separate analysis (see below). Both participants and items were included as two crossed random effects (Goldstein, 1995; Quené & Van den Bergh, 2008). Computations and evaluations were done with functions from the packages lme4 (Bates, Maechler, & Bolker, 2011) and languageR (Baayen, 2011; Baayen, Davidson, & Bates, 2008) for R (R Development Core Team, 2011).

The resulting coefficients and variances are listed in Table 3. Contrasts between response categories were

Table 3
Estimated parameters (*) for the mixed-effects regression of log-transformed response times (in log ms units). For fixed effects, regression coefficients are given, with probabilities based on MCMC simulation; for random effects, the standard deviations are given, with 95% confidence intervals based on MCMC simulation.

Fixed effects	Coefficient	(SE)	*t*	*p*
Exchanges (baseline)	6.566	0.026	252.7	<.0001
Anticipations	−0.079	0.030	−2.63	.0058
Perseverations	−0.153	0.050	−3.09	.0036
With additional error	+0.097	0.027	3.63	.0001
Random effects	Std. Dev.	95% C.I.	*N*	
Participants	0.1564	(0.0698, 0.1233)	78	
Items	0.0000	(0.0000, 0.0376)	56	
Residual	0.1932	(0.1954, 0.2365)	288	

Note: The reported estimates for the random effects may be unreliable, as suggested by the discrepancy between the estimates and their 95% confidence intervals based on MCMC simulation. The (distribution of the) random effects may be better captured by the median of MCMC simulated effects: between-participants $s = 0.0961$, between-items $s = 0.0071$, residual $s = 0.2154$.

evaluated by sign tests of the MCMC estimates corresponding to the appropriate contrasts. The first mixed-effects regression analysis shows a significant main effect of additional error (without additional error: mean 678 ms; with additional error: 786 ms; $p = .0001$ based on MCMC simulation). The two-way interaction between response category and additional error was not significant (a model including this interaction did not fit better, $p = .1766$). The regression coefficients in Table 3 confirm that, as predicted, response times for exchanges (mean 739 ms) are significantly slower than response times for anticipations (mean 674 ms, $p = .0044$ based on MCMC estimates) and than response times for perseverations (mean 649 ms, $p = .0016$). Response times for perseverations were only marginally faster than those for anticipations ($p = .0999$).

In addition, in the second mixed-effects regression analysis, the two fixed factors were, as before, response category (coded as dummy factors), and this time the absence or presence of a repair. Responses with and without additional error were pooled in this analysis. We have predicted that unrepaired errors have longer response times than repaired errors. This difference would be due to the way the speaker's self-monitoring deals with phonetic ambiguity. If the inner speech is ambiguous as to whether or not an error has been made, self-monitoring may be in an undecided state, which may lead to relatively slow responses. As in the previous analysis, participants and items were included as crossed random effects, and the same routines and packages were used. The resulting regression estimates are given in Table 4. The results of the second regression analysis of response times again show that these response times are significantly slower for exchanges than for anticipations and/or perseverations.

In addition, the second mixed-effects regression analysis shows a marginally significant main effect of repair (without repair: mean 716 ms, $n = 288$; with repair: 660 ms, $n = 34$; $p = .0950$ based on MCMC simulation). This shorter response time for repaired responses supports the idea that the inner speech inspected by the self-monitor is potentially ambiguous and varies in its degree of ambiguity. We will return to this issue in the discussion below. The two-way interaction between response category and repair was not significant (a model including this interaction did not fit better, $p = .1311$).

Of course a p value of .0950 is not very convincing. Fortunately, our prediction that unrepaired speech errors have longer response times than repaired speech errors could also be verified in the response times previously collected by Nooteboom and Quené (2008, Experiment 1&2 collapsed). There, in total 1343 errors were observed, of which only 2 were perseverations, 545 unrelated speech errors, 20 hesitations, and 257 omissions. These response categories were all excluded from regression analysis, with 515 errors remaining. The response times of these errors were again log-transformed and analyzed by means of the same mixed-effects regression as above; results are summarized in Table 5.

The resulting coefficients show a significant main effect of response category: responses containing an additional

Table 4
Estimated parameters (*) for the mixed-effects regression of log-transformed response times (in log ms units). For fixed effects, regression coefficients are given, with probabilities based on MCMC simulation; for random effects, the standard deviations are given, with 95% confidence intervals based on MCMC simulation.

Fixed effects	Coefficient	(SE)	*t*	*p*
Exchanges (baseline)	6.593	0.025	264.1	<.0001
Anticipations	−0.059	0.028	−2.10	.0330
Perseverations	−0.114	0.046	−3.09	.0222
With repair	−0.068	0.041	−1.67	.0950
Random effects	Std.Dev.	95% C.I.	*N*	
Participants	0.1691	(0.0791, 0.1288)	84	
Items	0.0134	(0.0000, 0.0394)	58	
Residual	0.1902	(0.1956, 0.2346)	322	

Note: The reported estimates for some random effects may be unreliable, as suggested by the discrepancy between the estimates and their 95% confidence intervals based on MCMC simulation. The (distribution of the) random effects may be better captured by the median of MCMC simulated effects: between-participants $s = 0.1029$, between-items $s = 0.0093$, residual $s = 0.2152$.

Table 5
Estimated parameters for the mixed-effects regression of log-transformed response times (in log ms units) from Nooteboom and Quené (2008, Experiment 1 and 2). For fixed effects, regression coefficients are given, with probabilities based on MCMC simulation; for random effects, the standard deviations are given, with 95% confidence intervals based on MCMC simulation.

Fixed effects	Coefficient	(SE)	*t*	*p*
Exchanges (baseline)	6.391	0.028	227.27	<.0001
Anticipations	−0.056	0.062	−0.91	.3136
Interruptions	+0.007	0.035	0.19	.9820
With additional error	+0.099	0.030	3.29	.0001
With repair	−0.111	0.030	−3.68	.0006
Random effects	Std.Dev.	95% C.I.	*N*	
Participants	0.1269	(0.0554,0.1122)	100	
Items	0.0579	(0.0000, 0.0715)	47	
Residual	0.2366	(0.2331, 0.2693)	515	

error yield a longer response time than exchanges and anticipations. This corresponds to the main effect of additional error in the above experiment (Table 3). In addition, the results show a significant main effect of repair, as predicted (without repair: mean 625 ms, with repair: mean 543 ms; $p = .0006$ based on MCMC simulation). The two-way interaction between response category and repair was again not significant (a model including this interaction did not fit better, $p = .213$).

Clearly, the distribution of exchanges, anticipations and perseverations over priming conditions in the current experiment is controlled by two major factors, first a strong tendency in the processes of planning and serial ordering of speech segments to favor exchanges over anticipations and perseverations and to favor anticipations over perseverations, and second an effect of priming for a specific error type. Of these two tendencies, the first is the most interesting because this is supposed to be caused by general properties of the processes of planning and serial ordering speech segments, whereas the second is specific to the experimental situation.

The analysis of response times confirms an earlier observation by Nooteboom and Quené (2008), demonstrating that making two consecutive errors, as in *barn door* turning into *dark boat*, during speech generation costs more time than making only the elicited speech error, such as *barn door* turning into *darn bore*. Supposedly, this difference in response times does not stem from the process of serial ordering, but rather from the process of self-monitoring. One thus would expect a similar difference between exchanges, each exchange for the monitor consisting of two consecutive errors, and anticipations or perseverations, each of which for the monitor consists of a single error. This is confirmed by the analysis, showing that exchanges have significantly longer response times than anticipations and perseverations. It was also predicted that on average repaired speech errors would have shorter response times than unrepaired speech errors. Also this expectation is corroborated by the data, and confirmed by an analysis of data obtained in earlier experiments.

Discussion

The results of the current experiment eliciting segmental exchanges, anticipations and perseverations with the SLIP technique lead to the conclusion that, at least in this experiment, the processes for planning and serial ordering of speech segments generate more segmental exchanges than anticipations and perseverations. This supports the claim by Nooteboom (2005a) that in inner speech exchanges are more frequent than anticipations and perseverations. The results of the investigation reported here favor a view of planning and serial ordering of speech segments that is qualitatively compatible with Shattuck-Hufnagel's scan-copier model (1979, 1983). The actual quantitative distribution of exchanges, anticipations and perseverations in this experiment, however, is not predicted by any of the current models of the serial ordering of segments during speech production. The reader will remember that Shattuck-Hufnagel assumed that an exchange results from a single slip in the process of serial ordering, viz. inserting a wrong, accidentally hyperactivated, segment like the *h* of *hemisphere* into the slot of another segment, thus for example turning *left hemisphere* into *heft.....* This is immediately followed by de-activation or inhibition of the wrongly inserted *h*, leaving the not inserted and therefore not de-activated *l* as the only reasonable candidate for the initial position of the second word, thus creating the exchange *heft lemisphere*. An anticipation would in this view result from two consecutive slips in the process, first inserting a wrong segment, and then not de-activating the inserted segment. A perseveration would result from first not de-activating a correctly inserted segment, and then re-inserting this segment again in the wrong position.

However, if one assumes that these two kinds of slips are independent, this process of serial ordering would never generate realistic distributions of the three types of segmental speech errors. In the absence of a more realistic model we therefore are forced to conclude that, where as each of the two kinds of slips in the process of serial ordering is relatively rare, perhaps in the order of once in a 1000 or 2000 words for spontaneous speech (cf. Levelt, 1989), as soon as such a slip has occurred the probability of the other kind of slip occurring in the same speech plan, increases enormously. Of course probabilities are different for the current experiment than for spontaneous speech due to the priming of speech errors and possibly other differences between the speaking conditions. In this experiment we obtained, over all three priming conditions, 228 exchanges, 80 anticipations, 24 perseverations and 74 early interruptions. These early interruptions are either half-way repaired exchanges or repaired anticipations. In order to estimate the numbers of segmental speech errors in inner speech, the 74 early interruptions have to be divided over exchanges and anticipations. A conservative way of doing this is to assume that the ratio between overt unrepaired exchanges and overt unrepaired anticipations reflects the ratio between exchanges and anticipations in inner speech. This is, given the purpose of this paper, conservative, because it assumes that the probability of being detected in inner speech is the same for exchanges and anticipations. Even so, following this reasoning we estimate that this experiment generated in inner speech some 280 exchanges, some 100 anticipations and 24 perseverations, demonstrating a strong predominance of exchanges.

Of course, these numbers very likely are strongly affected by the priming conditions, favoring exchanges over other segmental errors. It may be more realistic to repeat this exercise excluding the condition priming for exchanges. Doing this, we estimate that the experiment generated 109 exchanges, 68 anticipations and 21 perseverations, giving a ratio between the three types of segmental errors of roughly 5:3:1 for exchanges, anticipations and perseverations respectively. This comes satisfactorily close to the ratio between the three types of errors in spontaneous speech estimated from the study by Nooteboom (2005a), viz. 4:3:2. likewise for exchanges, anticipations and perseverations respectively. We conclude from these numbers that, if we assume a model similar to the scan-copier proposed by Shattuck-Hufnagel, we must also assume (a) a strong dependency between the two proposed consecutive slips in the process, i.e. inserting a wrong segment and not de-activating an inserted segment, and (b) that the strength of this dependency is sensitive to the order in which the two slips in the process occur, being stronger for the order generating anticipations than for the order generating perseverations. The dependency, of course, is such that, although each of the two kinds of slips in the process of serial ordering is relatively rare, as soon as one of these slips occurs in generating a mental plan for speaking, the probability that the other type of error occurs within the same mental plan increases enormously. Unfortunately, using the above ratios between error types for adapting the scan-copier model such that it generates more or less correct error frequencies will not provide any further insight. It would simply mean fitting the model to the data. It also should be acknowledged that computational models like the one proposed by Dell (1986) and WEAVER++ proposed by Levelt et al. (1999), although failing in explaining the apparent predominance of segmental exchanges in inner speech, in other respects account for a much wider range of phenomena than the rather abstract scan-copier model proposed by Shattuck-Hufnagel (1979, 1983). We suggest that ways should be explored to enrich the current computational models with a mechanism generating a predominance of exchanges in inner speech. Such a mechanism would probably be qualitatively similar to the relevant aspects of Shattuck-Hufnagel's scan-copier model.

The results of our experiment clearly show that if the number of interrupted errors is artificially reduced so that exchanges become less hidden in the category of interrupted errors, exchanges become easily the most frequent type of segmental errors. This is also clear evidence that in spontaneous speech the distribution of overt segmental errors over the three types exchanges, anticipations and perseverations is to a large extent controlled by the process of self-monitoring inner speech for speech errors, as has been assumed by many authors (Baars et al., 1975; Blackmer & Mitton, 1991; Hartsuiker, 2006; Hartsuiker, Kolk, et al., 2004; Levelt, 1983; Levelt, 1989; Levelt et al., 1999; Motley, 1980; Motley, Camden, & Baars, 1982; Nickels & Howard, 1995; Nooteboom, 2005a; Nooteboom, 2005b; Postma, 2000; Shattuck-Hufnagel, 1983).

The results on response times in the speech error elicitation experiment showed that segmental speech errors accompanied by an additional error have longer response times than simple segmental errors, confirming an earlier finding by Nooteboom and Quené (2008). These authors explained this from the extra time needed by the monitor to reject the simple error that was elicited in the experiment, and replace it with another error. In the current experiment it was also found that response times for unrepaired exchanges are (marginally) longer than those for unrepaired anticipations and perseverations. This can be explained by assuming that each suspicious segment, also when it is not classified as an error, requires some attention from the monitor, thereby increasing response time. One may note that apparently response times do not so much reflect the number of slips in the serial ordering process, but rather the number of segmental errors requiring attention during self-monitoring.

The results of our experiment also showed that unrepaired segmental errors have longer response times than repaired segmental errors. This was predicted from considerations related to the research by Goldstein et al. (2007), Pouplier and Goldstein (2005) and McMillan and Corley (2010), discussed in the introduction to this paper. This body of research suggests that in articulation there may be no all-or-nothing categorical speech errors, but only varying degrees of blending of competing articulatory gestures, possibly getting their degree of activation from (partly) activated competing speech segments "one level higher up" (cf. McMillan & Corley, 2010). We assume that this "one level higher up" corresponds to the form of inner speech that (1) is brought about and acted upon by the process of serial ordering segmental units (2) is the input for the process of self-monitoring for pre-articulatory speech errors, and (3) is the input for the articulation process. Given that articulation may demonstrate simultaneous activity of two competing articulatory gestures, we must assume that in speech preparation, due to errors in the process of serial ordering, segmental units of speech in inner speech can be simultaneously activated, competing for the same slot in the suprasegmental frame. Such ambiguity potentially causes response conflict to a monitoring process, and therefore is expected to increase reaction or response times (Botvinick et al., 2001; Szmalec et al., 2008). To phrase this differently, the probability of being detected and repaired would decrease and response times increase with increasing ambiguity of the error segment. From this hypothesized relation between probability of being detected and response times for error segments, one expects that unrepaired segmental errors on the average have an higher degree of ambiguity and therefore longer response times than repaired error segments. This is precisely what we found, albeit within the current experiment with only marginal significance. This finding was, however, strongly and significantly confirmed in analyzing data from earlier experiments, clearly and convincingly demonstrating that in a speech error elicitation experiment unrepaired errors have significantly longer response times than repaired speech errors. This supports the idea that in inner speech a segmental speech error can result in an ambiguous segment exhibiting properties of two competing segments. Admittedly, other explanations for this finding cannot be excluded. For example, fluctuations in vigilance might lead to greater arousal in some trials than in others, leading to both more effective monitoring and faster responding in the trials with accidentally higher vigilance, thus accounting for the observations without a need for assuming the blending of speech segments. Alternatively, fast responding might leave more time for repair than slow responding, causing the observed relation between response times and repairs. Obviously, the possible consequences of blending speech segments in segmental errors of speech for speech perception and for self-monitoring inner speech require further investigation. Such work is currently underway.

Conclusion

This paper is concerned with the frequency of exchanges relative to the frequencies of anticipations and perseverations in segmental speech errors. The results presented clearly show that in inner speech exchanges, before being filtered out by self-monitoring for speech errors, outnumber anticipations and perseverations. This can be explained by assuming a model for the serial ordering of segments during speech preparation that is qualitatively similar to Shattuck-Hufnagel's scan-copier model (Shattuck-Hufnagel, 1979, 1983), enriched such that once one of two kinds of slips occurs in the process of serial ordering, the probability that the other kind of slip will occur within the same mental programme increases enormously. Our finding that response times are longer for unrepaired than for repaired segmental speech errors supports a view of speech planning according to which two segments may, being completely or partially activated, simultaneously compete for the same slot in the suprasegmental framework, as suggested by results obtained by Goldstein et al. (2007) and McMillan and Corley (2010). Although other explanations are possible, we suggest that this difference stems from unrepaired speech errors on average being more ambiguous, and therefore requiring more attention from the monitor, than repaired speech errors. That in collections of speech errors in spontaneous speech complete exchanges are less frequent than anticipations and perseverations is also caused by the process of self-monitoring for speech errors: Most exchanges in inner speech are detected and repaired after the anticipatory part and before the perseveratory part of the exchange has been spoken. This causes the exchanges to be classified as repaired anticipations.

Acknowledgments

Our thanks are due to Theo Veenker for technical assistance and to Gary Dell and Rob Hartsuiker for sharing their thoughts on many aspects of the research reported here.

The raw data of the experiment are currently available on-line in the form of an excel document at the following web-adress: http://www.let.uu.nl/~Sieb.Nooteboom/personal/Experimentaldata.htm.

Appendix A

Stimulus word pairs used in the experiment. PRECURSORS gives the number of preceding word pairs, used for priming a particular segmental error (except in case of fillers). WORD PAIR gives the actual word pair to be spoken on cue by the participants. TYPE gives the type of segmental errors to be elicited: E = exchanges, A = Anticipations, P = Perseverations, F = Fillers. The list gives the stimuli as used for 48 odd-numbered participants. The list for the 48 even-numbered participants was identical, except that to-be-elicited anticipations were turned into to-be-elicited perseverations and vice versa by switching first and second words of the word pair, both in the stimuli and in the precursors.

Stimulus	Precursors	Word pair	Type	Group
1	5	paf kap	E	1
2	5	doos boot	E	1
3	5	voet zoel	E	1
4	5	vaal baak	E	1
5	5	baar vaal	E	1
6	5	vaan baas	E	1
7	5	vet zen	A	1
8	5	deed been	A	1
9	5	bos vod	A	1
10	5	kan pauw	A	1
11	5	bun pul	A	1
12	5	pon tof	A	1
13	5	duik buit	P	1
14	5	doop zooi	P	1
15	5	kaal taai	P	1
16	5	dom bok	P	1
17	5	pal kak	P	1
18	5	boen poet	P	1
19	5	keet pees	E	1
20	5	beuk peur	E	1
21	5	por tol	E	1
22	5	vil git	E	1
23	5	ban dal	E	1
24	5	kaap gaas	E	1
25	5	dik tip	A	1
26	5	gas kat	A	1
27	5	kool poon	A	1
28	5	del ben	A	1

Appendix A (*continued*)

Stimulus	Precursors	Word pair	Type	Group
29	5	veen zeem	A	1
30	5	pier kiep	A	1
31	5	zon dot	P	1
32	5	zaad vaag	P	1
33	5	keer teek	P	1
34	5	kot gom	P	1
35	5	gijn feit	P	1
36	5	buk pus	P	1
37	4	vaam tip	F	1
38	4	ros feil	F	1
39	4	vet pot	F	1
40	4	puim boef	F	1
41	3	wieg keus	F	1
42	3	maak juk	F	1
43	3	mom vies	F	1
44	3	dijn koor	F	1
45	2	git mik	F	1
46	2	big loot	F	1
47	2	wijn tuit	F	1
48	2	kir waag	F	1
49	2	heem lijp	F	1
50	2	ruik laaf	F	1
51	2	gif dep	F	1
52	2	ring koon	F	1
53	2	wijf ruig	F	1
54	2	kit waan	F	1
55	2	haan lijs	F	1
56	2	ruis heet	F	1
57	1	lof heg	F	1
58	1	nip hef	F	1
59	1	guit heit	F	1
60	1	duim hiel	F	1
61	1	rib wen	F	1
62	1	wak hel	F	1
63	1	loof haar	F	1
64	2	ruin lies	F	1
65	0	vim kil	F	1
66	0	woed looi	F	1
67	0	ris meel	F	1
68	0	moet neut	F	1
69	0	hoop laai	F	1
70	0	look haas	F	1
71	0	jaag hof	F	1
72	0	mik reeg	F	1
73	0	woef leen	F	1
74	0	ving kog	F	1
75	0	deur wies	F	1
76	0	deeg biet	F	1
77	0	baar lief	F	1
78	0	vaam kien	F	1
79	0	hos gup	F	1
80	0	hor weef	F	1
81	0	heil noor	F	1
82	0	riem dof	F	1

References

Baayen, R. H. (2011). *languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics"*. R package version 1.2. <http://CRAN.R-project.org/package=languageR>.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Bates, D., Maechler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-40. <http://CRAN.R-project.org/package=lme4>.

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, 3*, 28–47.

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

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

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychology Review, 108*(3), 624–652.

Cohen, A. (1966). Errors of speech and their implication of understanding the strategy of language users. *Zeitschrift für Phonetik, 21*(1/2), 177–181.

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

Dell, G. S., Ferreira, V. S., & Bock, K. (1999). Commentary on W.J.M. Levelt, A. Roelofs and A. Meyer: A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 41–42.

Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science, 17*, 149–195.

Efron, B., & Tibshirami, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Fromkin, V. A. (1993). The nonanomalous nature of anomalous utterances. *Language, 47*, 27–52.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). London: Edward Arnold.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition, 103*, 386–412.

Hartsuiker, R. J. (2006). Are speech error patterns affected by a monitoring bias? *Language and Cognitive Processes, 21*(7–8), 856–891.

Hartsuiker, R., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related Beply to Baars, Motley, and MacKay (1975). *Journal of Memory and Language, 52*, 58–70.

Hartsuiker, R. J., Kolk, H. H. J., & Martensen, H. (2005). Division of labor between internal and external speech monitoring. In R. Hartsuiker, Y. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 187–205). Hove: Psychology Press.

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

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

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

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–75.

McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition, 117*, 243–260.

Motley, M. T. (1980). Verification of "Freudian" slips and semantic prearticulatory editing via laboratory-induced spoonerisms. In: Fromkin, V. A. (Ed.), *Errors in linguistic performance. Slips of the tongue, ear, pen, and hand* (pp. 133–148). New York: Academic Press.

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

Nickels, L., & Howard, D. (1995). Phonological errors in aphasic naming: Comprehension, monitoring, and lexicality. *Cortex, 31*, 209–237.

Nooteboom, S. G. (1973). The tongue slips into patterns. In V. A. Fromkin (Ed.), *Speech Errors as Linguistic Evidence* (pp. 144–156). The Hague: Mouton.

Nooteboom, S. G. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand* (pp. 87–95). New York: Academic Press.

Nooteboom, S. G. (2005a). Listening to one-self: Monitoring speech production. In R. Hartsuiker, Y. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 167–186). Hove: Psychology Press.

Nooteboom, S. G. (2005b). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech Communication, 47*(1–2), 43–58.

Nooteboom, S. G., & Quené, H. (2008). Self-monitoring and feedback: A new attempt to find the main cause of lexical bias in phonological speech errors. *Journal of Memory and Language, 58*, 837–861.

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

Pouplier, M., & Goldstein, L. (2005). Asymmetries in the perception of speech production errors. *Journal of Phonetics, 33*, 47–75.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59*, 413–425.

R Development Core Team (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org/>.

Rossi, M., & Peter-Defare, E. (1998). *Les lapsus*. Paris: Presses Universitaires de France.

Schelvis, M. (1985). The collection, categorisation, storage and retrieval of spontaneous speech error material at the Institute of Phonetics, Utrecht. *PRIPU, 10*, 3–14.

Shao, J., & Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial order mechanism in sentence production. In Cooper W. E., & Walker, E. C. T. (Eds.), *Sentence processing* (pp. 295–342). Hillsdale, NJ: Erlbaum.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In MacNeilage, P. F. (Ed.), *The production of Speech* (pp. 109–136). New York: Springer.

Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior, 18*, 41–55.

Szmalec, A., Verbruggen, F., Vandierendonck, A., De Baene, W., Verguts, T., & Notebaert, W. (2008). *Neuroscience Letters, 435*, 158–162.

Van Alphen, P. M., 2004. *Perceptual relevance of prevoicing in Dutch*. Unpublished doctoral dissertation, Radboud University, Nijmegen, The Netherlands.

Venables & Ripley (2002). *Modern applied statistics with S*. New York: Springer.