Phonological similarity and lexical bias in phonological speech errors: self-monitoring or feedback?

Susannah Moat

MSc Psycholinguistics

*The University of Edinburgh*

2005
Abstract

The lexical bias effect refers to the fact that phonological errors result in real words more often than would be predicted by chance. It has also been observed that phonemes are more likely to be exchanged if they are phonologically similar. Both of these patterns of errors are easily explained within the framework of a feedback model (e.g. Dell, 1986), through feedback from phonemes to lexemes, and through feedback from features to phonemes. However, a feedforward account of these effects has also been proposed, which relies on a monitor mechanism that edits out nonwords and is less likely to reject segments similar to the intended utterance (Nooteboom, 2005).

Closer analysis reveals, however, that these two models make differing predictions concerning the interaction of phonological similarity with the lexical bias effect. The feedback model predicts that connectivity between the feedback loops concerned will result in mutual amplification of the two effects. Therefore, according to the feedback model, lexical bias will increase with phonological similarity. Conversely, Nooteboom (2005) postulates that errors resulting in real words are accepted as lexical by the self-monitor regardless of phonetic similarity, but nonword errors are less likely to be detected if they are phonetically similar to the intended utterance. The adapted monitor model therefore predicts that lexical bias will decrease with increasing phonological similarity.

This dissertation reports a speech error experiment using the Word Order Competition paradigm (Baars & Motley, 1976), in which phonological similarity and the lexicality of error outcomes are explicitly manipulated. The experiment replicates the lexical bias and phonological similarity effects, previously uninvestigated in this experimental paradigm, and the interaction uncovered between the two effects adds to the ever increasing pool of evidence for the existence of feedback in language production.

This work has been accepted as a poster at AMLaP 2005, Ghent, Belgium.
Contents

Introduction ......................................................................................................................2
Method.........................................................................................................................2
Results.........................................................................................................................2
Discussion ....................................................................................................................2
References ...................................................................................................................2
Acknowledgements .....................................................................................................2
Appendix 1: Materials..................................................................................................2
Modern theories of language production are by and large all based around the same representational levels, starting at the high level semantics of a word, and extending downwards through its syntactic (or lexical), phonological and phonetic representations towards final articulatory output (e.g., Dell, 1986; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999). However, as yet, no consensus has been reached over connectivity between these levels. Specifically, whilst all models propose that each level passes activation to the level below, the reciprocity of these connections is debated. Such reciprocity is known as feedback. Models such as Dell’s (1986) propose feedback between all levels of the language production system. However, others such as Levelt (1989; Levelt, Roelofs, & Meyer, 1999) take a feedforward approach, claiming that feedback is absent from the lexical-phonological connections downwards.

A result which has been important in considering this issue is the lexical bias effect. The lexical bias effect refers to the observation that phonological exchange errors with real word outcomes are more likely to occur than exchange errors with nonword outcomes. For instance, the error \textit{big pan} $\Rightarrow$ \textit{pig ban} would be more likely to occur than the error \textit{bald pal} $\Rightarrow$ \textit{pald bal}. This result is particularly compelling evidence for the feedback account, as it can easily be explained by observing that word outcome errors will be reinforced as a result of activation passing from the target word to its component phonemes, and subsequently travelling back up from the phonemes shared between the target and error, to the error outcome word representation. Nonwords cannot be reinforced in this manner, as by definition, they do not have lexical representations in the language production system. For instance, in the \textit{big pan} example, activation would pass from the lexical nodes \textit{BIG} and \textit{PAN} to /b/, /ɪ/ and /g/, /p/, /æ/ and /n/. The phonemes /ɪ/, /g/ and /p/ would then forward activation to \textit{PIG}, thereby increasing the chances that it should be produced in error. This would clearly not occur in the case of \textit{bald pal}, as neither the lexical node for \textit{pald}, nor the node for \textit{bal} exist.
Introduction

There has, however, been some disagreement over whether this effect is real, or whether chance could equally as well predict the patterns of speech error observed. Garrett (1976) conducted a corpus analysis in which he attempted to address this question, by estimating chance through sampling word pairs from published interviews, and exchanging their initial sounds. He found that 33% of these pseudo-errors created words. On the basis that this figure was not vastly different to the proportional of real word errors witnessed in natural speech, Garrett concluded that the lexical bias effect did not in fact exist. Similarly, Del Viso, Igoa, and García-Albea (1991) did not find evidence for the lexical bias effect in a corpus analysis of Spanish speech errors.

However, Nooteboom (2005) highlights that it is much more likely that a word will be created when the consonant in a monosyllabic word is changed, than in a polysyllabic word. As Garrett did not distinguish between monosyllabic and polysyllabic words in his estimation of chance, Nooteboom (2005) suggests that he may have obscured any effect of lexical bias by counting too many monosyllabic words. In a similar manner, the lack of lexical bias in Del Viso et al.’s (1991) study could be accounted for by the fact that Spanish contains a greater number of longer words than English (Berg, 1991), and so opportunities to create a new word in spontaneous speech will be decreased (Nooteboom, 2005). To support such suggestions, convincing evidence of the lexical bias has been found through a number of experimental studies (e.g., Baars, Motley, & MacKay, 1975; Dell, 1986, 1990; Hartsuiker, Corley, & Martensen, 2005; Humphreys, 2002; Nooteboom, in press). Given that chance can be controlled properly in such studies, and that other corpus studies such as Dell and Reich (1981) and Nooteboom (2005) in which chance was estimated in a more detailed manner have succeeded in uncovering a lexical bias effect, it seems safe and reasonable to conclude that the lexical bias effect does in fact exist.

Confirmation of this result is however not enough to assume the existence of feedback. The lexical bias effect can also be explained in feedforward models, which are equipped with a monitor mechanism (e.g. Levelt, 1989; Levelt et al., 1999). In the standard perceptual loop model (e.g., Hartsuiker & Kolk, 2001; Levelt, 1989; Levelt et al., 1999), speech is monitored both internally and externally, by feeding either an abstract phonological code or overt speech back into the speech comprehension system. At the conceptual level, the parsed speech is compared with the intended message, in order to
ascertain whether the output was correct. As part of this process, it is hypothesised that
the monitor checks utterances for lexicality. In this manner, it is proposed that more
nonword outcome errors (such as bald pal → pald bal) are removed than real word
outcome errors (such as big pan → pig ban), thereby creating a lexical bias effect.

This dissertation attempts to distinguish between the monitor and the feedback accounts
for lexical bias, by considering a second speech error pattern, the phonological similarity
effect. The phonological similarity effect is the result that phonological exchange errors
are more likely if the phonemes exchanged are similar (e.g. MacKay, 1973; Wilshire,
1999). For instance, the error look right → rook light would be more likely than the error
lace ford → face lord, as whereas /l/ shares its place of articulation feature (alveolar) and its
voicing feature (voiced) with /r/, it does not share any features with /f/.

Again, the feedback account explains the phonological similarity effect with ease.
Activation from planned phonemes will, as normal, pass to their phonological
subcomponents (often portrayed as features), which will in turn feed activation back to
phonemes which also use these features. In this way, phonemes which overlap in
representation with the target phoneme will be primed, whereas others will not. For
example, whereas activation of /l/ would also increase activation of /r/, as a result of
activation passing via their place of articulation and voicing features, activation of /l/
would not directly affect the level of activation of /f/.

The original perceptual loop monitor would not predict the phonological similarity
effect. However, Nooteboom (2005) has recently proposed a modification to the model
in order to allow it to do just that. He claims that, in contrast to the original monitor
model, the monitor has access to the target utterance, and carries out a phonological
comparison between each utterance output and its intended form. By this hypothesis, if
an error is phonologically similar to its target, it is more likely to be mistaken for the
target, and therefore not edited out. Therefore more errors will occur as phonological
similarity increases between the target and error, resulting in the phonological similarity
effect.

The motivation of the current experiment is the fact that Nooteboom’s model predicts a
different interaction between the lexical bias effect and the phonological similarity effect
than the feedback account would. As part of his modification to the monitor model, Nooteboom (2005, in press) specifies that word outcome errors are unlikely to be noticed by the monitor as they will simply be recognised as a word, regardless of phonological similarity. Nonword errors, however, will be less easily detected when they are more phonologically similar to the target utterance. Therefore, as nonword errors increase with phonological similarity, but the rate of word errors is not affected, the modified monitor model predicts that lexical bias decreases as phonological similarity increases.

However, the feedback model predicts the opposite interaction. This prediction results from the connectivity between the feedback loop which causes lexical bias and the feedback loop which causes phonological similarity. In the case of *look right* → *rook light* for example, activation would pass from /l/ to /r/ via the phonological similarity loop, and from there, would enter the lexical bias loop, giving extra activation to *rook*. The lexical bias effect would therefore be amplified. By this logic, the feedback model predicts that lexical bias should increase with phonological similarity.

To date, there have been three studies which have attempted to uncover the true form of the interaction between lexical bias and phonological similarity. One of these was a corpus study conducted by Nooteboom (2005). In this study, Nooteboom (2005) found that lexical bias decreased with phonological similarity, a result which would support his adapted monitor model. However, although Nooteboom attempted to overcome the usual limitation of corpus studies into lexical bias, by reducing the set of errors he considered to monosyllabic words, and then estimating chance by calculating how many words and nonwords could be created by replacing any phoneme in each word with another phoneme, there are still problems with these calculations. For instance, Nooteboom only considered the possibility of creating words or nonwords for the monosyllabic words in his error corpus. This ignores the large number of words which will not have occurred in the corpus as no errors resulted from them, making the sample population distinctly biased. In addition, the exact form of the interaction is somewhat obscured in Nooteboom’s analysis, as the overall number of errors in each phonological similarity band differed largely, and no attempt was made to estimate how much of this difference was caused by the phonological similarity effect, and how much was again the result of a sampling bias.
Nootenboom (in press) again considered this interaction using the SLIP task, an artificial error elicitation technique (Baars & Motley, 1974). This study was superior to his corpus analysis, in that the experimental setting allowed chance to be properly controlled. However, the pattern of lexical outcome errors that Nootenboom (in press) uncovered, such that there are many more errors in the medium phonological similarity lexical outcome condition than any other condition, was very surprising. Nootenboom explained this pattern as simply noise in the data. Such an explanation does not, however, render the remaining data entirely valid, thereby again providing few pointers as to the true form of the interaction.

The final study which has considered this question was a post-hoc examination of the pattern of errors found in Hartsuiker et al.’s (2005) SLIP task experiment into the lexical bias effect (McMillan, Corley, & Hartsuiker, 2005). McMillan et al. found the opposite result to Nootenboom (2005), namely that lexical bias increased with phonological similarity. This result would support the feedback model. However, Hartsuiker et al.’s (2005) experiment was intended to determine the modulation of the lexical bias effect by the context in which nonword stimuli were produced; i.e., whether real words were present in the stimulus list or not (c.f. Baars et al., 1975; Humphreys, 2002). Therefore, materials were not controlled for phonological similarity, resulting in the post-hoc analysis having only limited power.

The goal of the current experiment is therefore to uncover the form of this interaction, by conducting an experiment in which phonological similarity and lexical outcome are properly controlled. Given the simplicity of the feedback account in comparison to the somewhat ad-hoc nature of the modified monitor account, it is predicted that evidence in support of feedback will be observed, such that lexical bias will be found to increase with phonological similarity. Evidence of the opposite interaction, such that lexical bias decreases with increased phonological similarity, would provide support for Nootenboom’s (2005) adapted monitor model.

In order to do this, we will employ the Word Order Competition (WOC) task (Baars & Motley, 1976). In the WOC task, participants are rapidly presented with pairs of words or nonwords. These pairs are then replaced by arrows, pointing either left or right. The arrow direction signals the order in which the participant should produce the last word
pair if a beep should sound as the arrow is presented. If the arrow points right, participants should say the first word of the pair followed by the second word; i.e., the words should be produced in the same order as seen on the screen. If the arrow points left, participants should say the second word followed by the first word; i.e., the word pair should be reversed.

All targets are followed by a leftwards pointing arrow, and are preceded by 2 to 4 filler pairs, which are followed by rightwards pointing arrows. In this manner, participants are primed to produce the pairs from left to right, following the standard reading direction for English. However, on target pairs, the leftwards pointing arrow dictates that they must produce the pair from right to left, thereby causing an ordering confusion, which is intended to lead to exchange errors.

Most previous speech error elicitation experiments have not used this method, and have instead relied on the SLIP task (Baars & Motley, 1974). The SLIP task is similar to the WOC task, in that participants are again rapidly presented with pairs of words, which then disappear, and should only be produced if a beep should sound. The difference in the SLIP task is that error elicitation relies on phonological priming. For instance, a target pair such as *sup cub* → *cup sub* would be preceded by bias pairs such as *call sun* and *cuss suck*, in which the same onset consonants are used as in the target pair, but in the opposite order. This methodology has been estimated to generate exchange errors in around 6-10% of target utterances (Bock, 1996), although in recent experiments, the exchange error generation rate has only been between 2 and 5% (Hartsuiker et al, 2005; McMillan, 2004; Nooteboom, in press).

A myriad of results have been uncovered using this paradigm, from the effect of word frequency on exchange errors (Dell, 1990), to how an attractive experimenter can alter the phonological speech errors you make (Motley & Baars, 1979). However, questions must be raised over the ecological validity of the SLIP task. For instance, it has been suggested that errors may result from articulatory interference at the motor level (Baars & Motley, 1976). Participants are sometimes observed to subvocalise the filler items and bias words (Baars & Motley, 1976), which could mean that errors simply result from participants repeating the same articulatory motions. Articulation is clearly dealt with at a much later level than that at which words and phonemes are prepared, and so these
errors are unlikely to give the clearest picture of the aspects of the system we are interested in.

Another method sometimes used for exchange error generation is the tongue twister task (e.g., Wilshire, 1999). As the name would imply, this paradigm consists of giving participants tongue twisters to produce, which as in those used for amusement, causes a number of exchange errors to be elicited. However, again, concerns have been raised as to the level at which these errors are elicited, and whether the problem does not simply arise at the articulatory level as a result of the tongue making fast repeated similar movement (see Wilshire, 1999 for a review).

The WOC task avoids all of these criticisms, by relying on high level ordering confusion to elicit errors. The error patterns produced with this paradigm should therefore be maximally illuminating as to the architecture and mechanisms involved in the language production system. This paradigm has so far not been widely used since its creation in 1976 (Baars & Motley, 1976), probably as a result of researchers wishing to maintain comparability of results and therefore using the SLIP task. However, it was recently successfully employed by Pillon (1998) and Melinger (2003), to examine the role of morpheme units in speech production, by considering whether exchanges of morpheme units are more likely than exchanges of other multphoneme or syllable sized units.

One of the apparent advantages of the WOC task is its error rate. Baars and MacKay (1978) claim that the technique is very efficient at eliciting spoonerisms, generating errors on 20-40% of trials. However, examination of the literature does not appear to back this claim, revealing only 3.4% onset exchanges in Baars & Motley’s (1976) study, 3.8% exchange errors in Pillon’s (1998) study, and only 1.7% prefix exchanges in Melinger’s (2003) experiment. This study will therefore also have as a secondary aim to evaluate the efficiency of the WOC task as an onset exchange error generation paradigm.

Another methodological difference in this study will be the use of the Frisch (1996) measure of phonological similarity. Most previous studies into the phonological similarity effect have relied on feature counting (e.g., MacKay, 1973; Nooteboom, 2005, in press; Wilshire, 1999). However, this approach is deficient in a number of ways. Firstly, it seems somewhat simplistic to weight each feature equally. Secondly, to make
no differentiation between small changes in place or manner and large changes in place or manner would similarly appear to remove relevant information from the measure. Thirdly, Frisch (1996) proposes that the co-occurrence of certain feature combinations results in redundant information, such that information about a feature configuration varies in importance depending on the configuration of the other features. To account for this, Frisch (1996) applies a redundancy hierarchy (Frisch, Broe, & Pierrehumbert, 1995) in which relations between features are considered as well as the features themselves. Based on the validity studies of this measure performed by Frisch (1996), it is hoped that this approach will make our results clearer and more informative.
2

Method

Participants
Forty experimentally naïve students and non-students from the University of Edinburgh community took part in the experiment in return for payment of £5. Twenty-one of the participants were male, and 19 were female. They ranged from 18 to 26 years old, were all native speakers of English, and reported no history of speech or language problems.

Design
A 2 x 2 design was created, which manipulated two levels of phonological similarity (low and high) and two levels of outcome (lexical and nonlexical). Whilst it was desired to use two maximally different levels of phonological similarity, pairs in the phonological similarity bracket between the two levels were created for exploratory analysis. For each condition, 20 pairs were created. Pairs were matched across the lexical and nonlexical conditions. Counterbalancing of materials across participants is explained under List construction.

Materials
It has been shown that low frequency words are more likely to be involved in phonological exchange errors than high frequency words (Dell, 1990). This could be explained through the production of lower frequency words being less reinforced in the language system than the production of higher frequency words. It would follow, therefore, that nonwords would be most likely to be involved in exchange errors.

On this assumption, we constructed 60 nonword pairs of the form CVC(C) CVC(C), such that the exchange of the initial consonants resulted in a pair of phonologically and orthographically valid words. The BEEP phonological dictionary (Robinson, n.d.), and
the ARC nonword dictionary (Rastle, Harrington, & Coltheart, 2002) were used to verify that the pronunciation of the orthographic exchange was the same as the outcome of swapping the onset phonemes of the target stimuli.

Targets were distributed evenly across 3 bands of phonological similarity, as determined by Frisch’s (1996) phonological similarity classification, resulting in 20 pairs in each phonological similarity band. Frisch’s calculations allocate pairs of consonants similarity values between 0 and 1, where 0 means that the consonants are not at all similar, and 1 means that the consonants are the same. To choose the boundaries between bands, the overall distribution of consonant pairs across similarity values was examined. Excluding pairs in which the consonants are the same (for which the similarity value is 1), the maximum value allocated is 0.58, and the minimum value is 0.03. However, there are many more pairs at the lower end of the distribution, such that the median similarity value is 0.13. Therefore, bands were selected such that a third of all possible non-identical consonant pairs would fall into each band, leaving the lowest band with an upper similarity value of 0.09, and the middle band with an upper similarity value of 0.18.

For each target pair, we constructed another nonword pair with a nonword outcome. Each nonword outcome pair was matched to the original lexical outcome pair in the onset and vowels used. For instance, for the lexical outcome pair gox fort → fox got, the nonlexical outcome pair gom fol → fam gol was created. This resulted in 120 experimental items, a full list of which is given in Appendix 1. Just as lexical outcomes were phonologically and orthographically valid, nonlexical outcomes were both phonologically and orthographically invalid. Stimulus words in the nonlexical outcome pair only differed in their codas from the stimulus words in the lexical outcome pair.

Nonlexical outcome pair equivalents for lexical outcome pairs were chosen based on a number of criteria. Firstly, the nonwords used in the stimulus pairs were picked from the same pool of nonwords used for the lexical outcome stimuli; i.e., nonwords formed by changing the first consonant of an existing word. Secondly, nonwords with as high a phonological neighbourhood as possible were selected, as initial material generations were resulting in nonlexical outcome pairs having lower phonological neighbourhood counts for higher phonological similarity ratings. Given evidence which implies that the sparser a phonological neighbourhood, the more likely an exchange error is (e.g.,
Vitevitch, 2002), it seemed sensible to avoid any possible confound of phonological
neighbourhood with phonological similarity. Choosing nonlexical outcome pairs with
denser phonological neighbourhoods succeeded in removing this confound. Thirdly, the
length of each nonword was kept as close as possible to the length of the matching
nonword in the lexical outcome pair, in order to reduce orthographic differences
between the pairs. Lastly, nonwords with as dense an orthographic neighbourhood as
possible were chosen, as in general, lexical outcome pairs had denser orthographic
neighbourhoods, and it was seen as desirable to reduce this difference between the two
conditions as far as was possible.

Materials were automatically selected, and four pairs were adjusted by hand after
inspection. Across the stimuli and outcomes, no words or nonwords, or their
pronunciations, were reused at all, in order to avoid unwanted priming which could bias
the likelihood of exchanges.

Constraints on the stimuli
As the experiment was conducted in Edinburgh, words containing any of the
orthographs er, ear, ir, or, or, ur, or beginning with wh, were rejected to avoid problems
resulting from Scottish pronunciation differing from standard English pronunciation.
Further constraints on the materials resulted from the planned usage of the same
materials in an ultrasound experiment, in order that articulation patterns during speech
errors could be examined more closely. For instance, the onset of one word was never
the same as, nor the voicing equivalent of, the coda of that word or the other word in the
pair. This was to make it easier to separate articulation patterns of the coda from
articulation patterns of the following onset phoneme. In addition, only low vowels and
diphthongs (/a/, /ɔ/, /ɛ/, /æ/, /ʌ/, /ʊ/, /eɪ/, /æɪ/, /ɜ/, /əʊ/ and /ɒ/) were used,
so that the tongue would make clear movements from the roof of the mouth between
consonants. Finally, as a result of the nature of the WOC task, no stimulus pair in the
experiment used codas which were voicing equivalents, so that word order and exchange
errors could be reliably detected, as is explained in more detail under Coding of responses.
Control of the stimuli

As a number of factors other than phonological similarity of onset consonants and lexicality of outcome are known to affect the likelihood of an exchange error, the materials were carefully controlled to avoid any unintended confounds. For instance, it has been shown that consonants which are followed by the same vowel in each word of the pair (e.g. *lace fate* → *face late*) are more likely to be exchanged than consonants which are not followed by the same vowel (e.g., *lace ford* → *face lord*) (Dell, 1984, 1986). Therefore, to remove this confound and maximise the chance of onset exchange error generation, all stimulus pairs used the same vowel in each word.

Secondly, it was noted that Pillon (1998) and Melinger (2003) both demonstrated in previous WOC studies that morphemes are more likely to be swapped than other phonemes or phoneme clusters. Therefore, all stimuli and outcomes were monomorphemic.

Thirdly, word frequency was considered. Frequency clearly affects exchange error processes, as it has been shown that low-frequency words are more likely to be involved in exchange errors than high-frequency ones (e.g., Dell, 1990). This result does not directly affect our materials, as all stimuli were nonwords. However, it does raise the possibility that the frequency of word outcomes may affect the probability of an exchange error, with exchange errors perhaps being more likely as frequency increases, due to strengthened connections for the lexeme unit. If this is the case, it would be preferable to avoid low-frequency outcomes, and so no word which was not listed in Kucera and Francis (1967) was used. It was also verified that word frequency of lexical outcomes was not significantly different across conditions, according to the Kucera-Francis frequency ratings (Kucera & Francis, 1967), obtained from the MRC database (Coltheart, 1981; Wilson, 1988).

Fourthly, Levitt and Healy (1985) suggested that less frequent phonological segments are more likely to be subject to error than more frequent phonological segments. It was therefore ensured that onset consonant frequency was not significantly different across conditions.
Fifthly, as was earlier noted, the phonological neighbourhood size of a word is hypothesised to affect the probability of an exchange error, with sparser neighbourhood sizes leading to higher incidence rates of onset exchange errors (e.g., Vitevitch, 2002). Therefore, it was confirmed that there was no significant difference in phonological neighbourhood size across conditions.

Nevertheless, the orthographic neighbourhood of stimuli was significantly higher for those in the lexical outcome condition than for those in the nonlexical outcome condition, despite efforts to reduce this difference. This was to some extent expected, as nonwords with larger orthographic neighbourhoods have more chance of being successfully paired with another word, such that the onset consonants can be exchanged and result in a lexical outcome. However, any effect which would be expected if phonological neighbourhood result patterns were to be extended to orthographic neighbourhoods would work against the lexical bias effect, as the nonwords in the nonlexical outcome condition would be expected to be involved in more exchanges as a result. This difference was therefore not too much of a concern.

**Filler pairs**

In addition to these target pairs, 230 filler pairs were developed. Two hundred and eighteen of these filler pairs were to be used as the biasing forwards pairs in the WOC task, in contrast to the target pairs which were to be said backwards. As it has previously been shown that a lexical bias effect is not observed if only nonwords are perceived (e.g., Baars et al., 1975; Hartsuiker et al., 2005), the filler pairs were 50% word-word pairs, and 50% nonword-nonword pairs, to avoid an entirely nonword context without making the nonword experimental pairs instantly identifiable.

The remaining twelve pairs were to be said backwards. Two were used as practice pairs at the beginning of the experiment to allow for the participant to acclimatise to the task, one of which was a word-word pair and the other of which was a nonword-nonword pair. The other ten were word-word pair foils that were added to break the pattern of all backwards items being nonword-nonword pairs, which could have made it easier for the participant to predict the direction in which the target word pairs were to be said.
All filler pairs had nonlexical outcomes if their onsets were exchanged. Neither the filler pairs nor their exchange outcomes were the same as any nonwords used as stimuli, nor the exchange outcome of any experimental pairs. Filler items were kept as similar to target items as possible in characteristics such as length, number of phonemes and neighbourhood size, again in order to avoid the participant predicting when backward arrows would appear.

Non-linguistic materials

In addition, a one minute sample of brown noise was created using Adobe Audition, to be played on repeat throughout the test period, through Sony Digital Reference Dynamic Stereo Headphones. Brown noise is similar to white noise, but rather than having an equal representation of all frequencies, lower frequencies are represented more than higher ones. Brown noise is therefore more pleasant to the human ear, as the ear perceives higher frequencies as much louder than lower frequencies. The brown noise was played to participants in order to reduce any possible monitoring of overt speech, in accordance with the approach of other speech error researchers such as Baars et al. (1975) and Hartsuiker et al. (2005),

List construction

Two experimental lists were created. To do this, ten pairs from each condition were randomly selected for use in one list, and their lexical or nonlexical equivalent was used in the other list. There were therefore 60 target pairs in each list, as well as the two practice items and the ten lexical foils. Each of the targets, practice items and foils was preceded by two to four filler items, resulting in a total of 290 pairs in each list. Filler items did not have the same onset phonemes as the target pairs in order to remove any phonological bias, as would be seen in a SLIP task. Filler items were the same in each list, so that lexical and nonlexical equivalents were preceded by the same fillers.

Four different orders of the two lists were used, in order to equalise fatigue effects across conditions. Each participant was presented with one order of one of the lists, resulting in 8 counterbalanced list conditions overall.
**Procedure**

The experimental procedure was largely based on the WOC paradigm, developed by Baars and Motley (1976), and also used by Pillon (1998) and Melinger (2003). Participants were tested individually in a quiet room, whilst listening to the previously described computer-generated brown noise through headphones. They were presented with word pairs, all in lower case, which after 1000ms were replaced by an arrow which was always accompanied by a beep. This arrow pointed either left or right. When the arrow pointed right, participants were asked to produce the first word in the pair followed by the second word; i.e., to repeat the words as they had seen them. When the arrow pointed left, participants were asked to produce the second word in the pair followed by the first word, thereby reversing the order of the words. Target items, practice items and lexical foils were always followed by arrows pointing left. All other fillers were followed by arrows pointing right.

The procedure used differed slightly to that of Baars and Motley (1976), as participants produced every pair, rather than being prompted to respond only for certain pairs. In the SLIP task which the WOC task is based on, it is crucial that filler pairs are not produced, in order to reduce the articulatory bias created by the phonological bias pairs. However, no such bias exists in the filler pairs of the WOC task, and it was noted that asking participants to produce all pairs would reduce the number of fillers to be created and to be processed during the experiment. This would reduce experimental time, transcription time, material generation time and recording tape required. After verification through a pilot study that errors were still being generated using this technique, this approach was adopted.

Naming latencies were recorded using a voice-activated microphone. If participants did not initiate a response within 1000ms, a loud buzzer sounded and the screen turned red for 250ms, to encourage them to speak faster on future trials. The next word pair appeared 400ms after response onset was recorded or the warning screen disappeared. Speed as well as volume of response was emphasised at all times, both in the instructions and also by telling participants that speed of response would be measured, in order to focus them on this issue.
Participants took part in a practice session before completing the actual experimental session, to ensure that instructions had been understood. The actual experiment took 10 minutes, with a one minute break in the middle, and required approximately 20 minutes total participant time including setting up, paperwork and the practice session.

**Coding of responses**

Responses were recorded on to Digital Audio Tape at 32,000Hz using a Sony TCD-D8 Digital Audio Tape-Corder and a Sony ECM-TS125 condenser microphone, and transcribed and coded offline. Following Pillon’s (1998) logic that participants intend to reverse the order of the stimulus pair, productions were scored with respect to the reversed outcome. Here, the reversed outcome is therefore specified for comparison in order to make error classification easier to understand. The categories used were: correct (such as *pung lulse → pung lulse*); full onset exchanges (such as *pung lulse → lung lulse*); partial onset exchanges, split into onset anticipations (such as *pung lulse → lung lulse*) and onset perseverations (such as *pung lulse → pung pulse*); failure to respond (silence, a noise not interpretable as an utterance or a response such as “don’t know”); and other error types, noted for information about the WOC task and for use during later analysis, including full rime exchanges (such as *pung lulse → pulse lung*); partial rime exchanges, split into rime anticipations (such as *pung lulse → pulse lulse*); rime perseverations (such as *pung lulse → pung lung*); and finally reversals (such as *pung lulse → lulse pung*).

Error coding was complicated by the possibility of reversed utterances, as errors such as *pung lulse → lung lulse* could have either been full onset exchanges, or reversals and full rime exchanges. As the influences on swapping each part of a word are likely to be very different, correctly distinguishing between the two was important. In cases such as the previous example, errors were coded as onset exchanges, based on the default assumption that the correct order had been intended. For example, given an outcome such as *pung lulse → lung lulse*, it was reasoned that the onset and rime of the second word and the rime of the first word were in the correct position for the non-reversed order, whereas only the onset of the first word was in the correct position for the reversed order, and that this error should therefore be classed as an onset anticipation error. Similarly, given an outcome such as *pung lulse → lulse lung*, it was reasoned that that the onset and rime of the first word and the rime of the second word were in the correct
position for the reversed order, whereas only the onset of the second word was in the correct position for the non-reversed order, and that this error should therefore be classed as a reversal and an onset perseveration error.
3

Results

Time to respond
The mean time to respond (excluding items for which the microphone did not trigger, and those for which participants took longer than 1000ms to respond) was 678ms. There were no differences in response latencies across conditions (all Fs < 2).

Mispronunciation removal
It was found that a number of the experimental items were prone to pronunciation which differed from the intended pronunciation. For instance, experimental pairs containing the letters “ou” were often pronounced as /u/ instead of the /ou/ which was predicted by the BEEP dictionary (Robinson, n.d.) and the experimenter. This meant that lexical outcomes were no longer lexical. Therefore, pairs containing this letter pair were excluded from analysis. In addition, it was noted that the onset consonant of eelp in eelp bent, and cest in the nonlexical matched pair cest heff was not consistently pronounced as /s/, and was often instead produced as /k/. This pair of items was therefore also excluded from analysis.

Response category
After mispronunciation removal, 1520 of the 1600 experimental trials remained. 582 of these trials were correctly named (38.29%), and there were 11 full onset exchange errors (0.72%), 33 partial onset exchange errors (2.17%). 894 other errors were recorded (58.81%) including 68 full rime exchanges (4.47%), 71 partial rime exchanges (4.67%), 149 reversal errors (9.8%) (8 of which were also partial onset exchanges, 4 of which were also partial rime exchange errors, and 1 of which was also a full rime exchange error), 42 failures to respond (2.76%), and 577 remaining errors (37.96%).

PHONOLOGICAL SIMILARITY AND LEXICAL BIAS
Only the phonological similarity of the onsets of words was controlled. Therefore, only onset exchange errors were included in analysis. The mean number of onset exchange errors in each experimental condition, weighted to account for mispronunciation removal is listed in Table 1.

Table 1: Mean number of full and partial onset exchanges in each condition of the experiment per participant, weighted to account for mispronunciation removal

<table>
<thead>
<tr>
<th>Phonological similarity</th>
<th>Lexical outcome</th>
<th>Nonlexical outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full</td>
<td>Partial</td>
</tr>
<tr>
<td>Low</td>
<td>0.075</td>
<td>0.05</td>
</tr>
<tr>
<td>High</td>
<td>0.188</td>
<td>0.519</td>
</tr>
</tbody>
</table>

Due to the low number of recorded full exchanges, we conflated full and partial (anticipation and perseveration) exchanges for all analyses reported. Analysis of full exchanges produced patterns which do not deviate significantly from the reported findings. A graph showing the mean number of (full and partial) onset exchange errors in each condition is given in Figure 1.

An ANOVA on the total number of onset exchange errors (full and partial combined) with outcome (lexical vs. nonlexical) as a within-subject and within-item variable, and phonological similarity (low vs. high) \(^1\) as a within-subject and between-item variable, showed a significant effect of outcome by subjects, and a marginally significant effect of outcome by items, such that lexical outcome errors were more likely than nonlexical outcome errors, thereby replicating the lexical bias effect \([F(1,39) = 7.674, \ p < .01; \ F(1,36) = 4.529, \ p = .065]\). It also revealed a significant effect of phonological similarity by subjects and by items, such that pairs in which the consonants were highly phonologically similar were more likely to be subject to exchange errors than pairs in which the consonants were phonologically dissimilar \([F(1,39) = 24.466, \ p < .0001; \ F(1,36) = 6.822, \ p < .02]\). Post-hoc analyses showed that there was a phonological similarity effect in both the lexical outcome condition \([t(39)=4.214; \ p < .0001]\) and the nonlexical outcome condition \([t(39)=2.769; \ p < .01]\).

\(^1\) As explained in the design section, the middle phonological similarity band was excluded from the main analysis, but exploratory analysis including this band did not deviate significantly from the analysis reported.
Finally, a marginally significant Outcome × Phonological Similarity interaction by subjects was observed [$F_1(1,39) = 3.001$, $p = .091$; $F_2(1,36) = 1.261$, $p = .269$]. Post-hoc analyses revealed that the lexical bias effect for low phonological similarity pairs was only marginal [$t_1(39) = 1.669$, $p = .052$] whereas for the high phonological similarity pairs, the lexical bias effect was strong [$t_1(39) = 2.411$, $p < .02$]. These findings hint towards the lexical bias effect being modulated by phonological similarity, such that lexical bias increases with phonological similarity.

![Figure 1: Mean number of (full and partial) onset exchange errors made per participant, by phonological similarity](image-url)
Discussion

The reported experiment replicated the lexical bias effect (e.g., Baars et al., 1975) and the phonological similarity effect (e.g., Wilshire, 1999) using the Word Order Competition (WOC) task, an experimental error elicitation paradigm which has not previously been used to examine these questions. Crucially, a marginally significant trend towards lexical bias increasing with phonological similarity was also observed.

The experiment therefore did not replicate the decrease of lexical bias with increasing phonological similarity which Nooteboom claimed that his corpus study results (Nooteboom, 2005) and SLIP task results (Nooteboom, in press) showed. Such a pattern would be predicted by Nooteboom’s adapted monitor model, in which a phonological comparison between target form and error causes the number of nonword errors to increase with phonological similarity, but does not affect real words, which pass by the monitor as a result of their lexical status. To the contrary, the results show that both lexical and nonlexical outcome errors increase with phonological similarity, and suggest the opposite interaction between the lexical bias and phonological similarity effect. This outcome, added to the problems with estimation of chance in Nooteboom’s (2005) corpus study, and the strange pattern of Nooteboom’s (in press) SLIP task results which he himself ascribed to noise in the data, leaves little support for Nooteboom’s adapted monitor model (Nooteboom 2005, in press).

However, the marginal significance of the interaction does not rule out the possibility that there is simply no interaction between the phonological similarity and lexical bias effect. Clearly, this outcome would not be predicted by the feedback account. Is it then possible that the monitor account could be altered to explain such a result?

The clear obstacle for Nooteboom’s (2005, in press) model in predicting no interaction between the two effects is the assumed superiority of the lexicality check over the
phonological comparison, such that real word outcome errors are likely to be accepted regardless of phonological similarity. However, on closer examination, it appears that this assumption does not make sense.

Firstly, Nooteboom (2005) himself has challenged the standard monitor claim that real word phonological errors are treated differently to nonword phonological errors as a result of their lexical status (e.g. Levelt, 1989; Levelt et al., 1999). He refers to previous research, demonstrating that the distribution of the number of words a speaker goes on speaking before stopping to correct a speech error, as well as the distribution of the number of words a speaker retraces in self correction, is significantly different for phonological errors in comparison to lexical errors, errors in which the speaker chose the wrong word for non-phonological reasons (Nooteboom, 1980). As the monitor is proposed to be responsible for error detection and correction, it can be assumed that these results should reflect the mechanisms and strategies which the monitor uses.

Analysis of Nooteboom’s current speech error corpus reveals that these distributions are not significantly different between real word phonological errors and nonword phonological errors, whereas significant differences are found when comparing lexical errors with real word phonological errors (Nooteboom, 2005). This finding is supported by results provided by Shattuck-Hufnagel and Cutler (1999), who showed that whereas lexical errors are corrected with a pitch accent on the corrected item, nonword and, most importantly, real word outcome phonological errors are not.

Nooteboom (2005) therefore concludes that, contrary to standard monitor theories, the monitor does not simply determine whether an utterance constitutes a word or not, but treats real word outcome phonological errors in the same way as nonword phonological errors. However, this conclusion seems to directly contradict his assumption that real word outcome phonological errors pass by the monitor as a result of their lexical status. To the contrary, his results would imply that within his model, real word outcome errors should be subject to the same phonological comparison error check.

Secondly, it is in any case not obvious how passing the lexicality check would exempt an utterance from the phonological similarity check. It would make more sense to assume that any given check would signal an error, rather than signalling a correct utterance.
If the lexicality criterion is therefore assumed not to be superior, it seems that given the current level of specification of the adapted monitor mechanism (Nooeboom, 2005, in press), it should not instantly be ruled out that the monitor model could be construed to account for a result pattern in which there was no interaction between the lexical bias effect and the phonological similarity effect. However, on closer examination, there is a large problem with this account. Specifically, it appears that the underlying explanation of how the monitor model can account for the phonological similarity effect is somewhat flawed.

Nooeboom (2005) argues that, contrary to standard perceptual loop monitor theories (e.g. Levelt, 1989; Levelt et al., 1999), the monitor has access to the target utterance, against which a phonological comparison check is carried out. However, this is not a change which can be implemented easily. The reason for Levelt et al.’s presumption that the monitor does not in fact have access to the target utterance originates deep in the basic architecture of the perceptual loop monitor. According to this theory, speech is monitored through an inner phonological code being fed back through the speech comprehension system, and a comparison being made at the conceptual level between the intended message and the parsed version of the monitored message (Hartsuiker & Kolk, 2001; Levelt, 1989; Levelt et al., 1999). This proposal permits Levelt and others to defend the monitor model against claims that it is overpowered and underspecified (e.g., Dell & Reich, 1980; Rapp & Goldrick, 2000) through maintaining that the monitor model is in fact very parsimonious, as it relies almost exclusively on sections and paths in the language system which already exist (e.g., Roelofs, 2004). Allowing the monitor to have access to the language production system and the phonological form of the intended utterance, such that this could be compared with the incoming utterance, clearly complicates the monitor mechanism substantially and therefore breaks this parsimony. Additionally, given that in an error, the wrong phonological form was clearly selected at some point during production, one must ask which level of production representation the monitor would refer to, in order to ensure a correct target for comparison.

If an adaptation of the standard perceptual loop monitor cannot account for the phonological similarity effect, could a monitor internal to the production system fill this gap instead? It has been suggested by a number of researchers that the proposed monitor system most likely consists of a combination of production and perception...
monitors (e.g., Nickels & Howard, 1995; Postma, 2000; Postma & Oomen, 2005). However, it is obvious that a production monitor cannot simply have a copy of an idealised output code which it uses to compare with the actual output code; otherwise, it would not be clear why the production system did not use this correct code in the first place (e.g., MacKay, 1987; Postma, 2000).

This leaves one feasible possibility as to how a phonological comparison would take place in a monitoring system; namely, that a production monitor compares the pattern of activation emitted from a level, with the pattern of activation returned to that level, and reacts if the feedback pattern is incorrect (e.g., Postma & Kolk, 1993; Vigliocco & Hartsuiker, 2002). By this logic, we have again arrived at the conclusion that feedback is present in the language production system.

It therefore appears that the existence of feedback is the only logical deduction. This leaves the question of why we did not find a more strongly significant interaction between lexical bias and phonological similarity. It may be the case that the experiment was not powerful enough. A plausible cause of such a problem could be the low amount of data generated by the WOC task, a problem which is discussed in more detail later. A second issue could be related to timing. Dell (1986) showed that no lexical bias is evident when a deadline of 500ms is used, and that a longer deadline of 700ms or 1000ms is required for the effect to become evident. Presumably, an interactive effect created by two feedback loops would require even longer to become detectable. Given that the current experiment used a speaking deadline of 1000ms, it is possible that at this interval, the effect was still too small.

The hypothesis of the existence of feedback is well supported by the number of results in the literature which this simple assumption very easily explains. Two of the most well-known are the mixed-error effect and the syntactic category effect. The mixed-error effect is the result that semantic substitution errors are more likely when the error outcome also shares phonemes with the target; e.g., CAT would be more likely to be replaced with RAT than with MOUSE. This effect has been widely demonstrated in corpus analyses (e.g. Dell & Reich, 1981; Harley, 1984), and in experimental settings (Ferreira & Griffin, 2003; Martin, Gagnon, Schwartz, Dell, & Saffran, 1996; Martin, Weisberg, & Saffran, 1989). A feedback model would predict exactly this effect, through an interaction of misactivation.
from higher semantic levels, and feedback from lower phonological levels converging on the lexical level. It has also been observed that phonological errors often share the same syntactic category as the target word (e.g., Fay & Cutler, 1977). This is the syntactic-category effect. Again, a feedback model would predict this effect as a result of the interactivity of the lexical level, where syntax is important, with the phonological form level.

 Whilst it has sometimes been claimed that there is no reason for feedback to exist other than for it to explain error patterns (e.g., Levelt et al., 1999), it is clear that the presence of feedback could in fact increase the efficiency and accuracy of the language production system. Postma (2000) and Vigliocco and Hartsuiker (2002) highlight that feedback may minimise the effects of noise, through feedback from lower levels reinforcing the correctly selected entries at higher levels and vice versa; for instance, the feedback loop between the lexeme CAT and the phonemes /k/, /æ/ and /t/ should help ensure that both the correct lexeme and the correct phonemes are activated sufficiently to be selected. Such an architecture would also help remove the need for the overcomplicated and very abstractly defined binding-by-checking mechanism required in Levelt et al.’s (1999) model, to ensure that phonemes selected match the intended lexical output (c.f. Dell, Ferreira, & Bock, 1999; O’Seaghdha, 1999), especially given that this mechanism is deficient in emulating known speech error patterns (Levelt et al., 1999) due to its overperfectionist nature.

 Furthermore, Vigliocco and Hartsuiker (2002) highlight the various roles of feedback in neural processing, and conclude that on neurological grounds, it is very unlikely that feedback does not exist. For example, they cite Spitzer’s (1999) statistic that only 0.1% of pyramidal cells in the cortex are involved in input and output operations. The rest are only concerned with internal connections. Vigliocco and Hartsuiker (2002) note that given the immense number of neurons and the fact that a pyramidal neuron has on average about 10,000 output connections, a simple calculation will show that a signal returns to where it started after an average of only three synaptic transmissions.

 On the basis of all this evidence, feedback in language production appears to be a very parsimonious hypothesis. However, whilst it does appear that monitoring without feedback cannot account for all the effects observed in this experiment, it should be
noted that such a conclusion should not be interpreted as a claim that monitoring does not occur at all in the language production system. To the contrary, it is currently unclear how evidence such as that provided by Motley, Camden, and Baars (1982), who showed that speakers are less likely to produce speech errors which result in taboo words than speech errors which result in neutral words, could be explained without the hypothesis of inner monitoring processes.

Furthermore, Hartsuiker et al. (2005) provide evidence which, once more, suggests that the lexical bias effect is indeed affected by a self-monitor. They show that the lexical bias effect is clear in a mixed context, where fillers in the stimulus list consist of both words and nonwords, but is absent in a nonword context, where participants perceive only nonwords. It is not clear how context modulation could be explained in a pure feedback model. However, it should be noted that the number of nonword errors does not differ significantly between context conditions. Rather, it is the number of word errors which is reduced in the nonword context. The standard monitoring account (c.f. Baars et al., 1975), which specifies that nonwords are edited out in a mixed context but not in a nonword context also does not predict this pattern. According to this model, it is the number of word errors which should remain the same between conditions.

Hartsuiker et al. (2005) proposed that this pattern of results could be explained by postulating a combination of both monitoring processes and feedback. They suggested that feedback causes an underlying lexical bias. However, in a nonword context, the self monitor removes more word outcome errors, as for the context, they are anomalous. This therefore removes lexical bias in a nonword context. In a mixed context however, lexical bias is not affected by monitoring, as the monitor has no criterion by which to remove errors. By supporting the hypothesis of the presence of feedback in the language production system, the current results are not inconsistent with Hartsuiker et al.’s (2005) model.

The WOC task

A secondary aim of this experiment was to evaluate the usefulness of the Word Order Competition (WOC) task (Baars & Motley, 1976) as an error elicitation technique. That the lexical bias and phonological similarity effects were replicated not only provides
support for the existence of these effects, but helps determine that results from the WOC task can be validly compared with previous speech error experiments. In addition, experimental time was shorter and stimulus generation less complex, and the fact that the data was known to be free of the biases of the SLIP task was very advantageous.

However, the task did also have some disadvantages. For instance, error coding was harder, as a result of the possibility that participants would fail to reverse the target as instructed. Indeed, our data shows that this occurred on 9.8% of trials. This problem made it necessary to use heuristics to determine the intended order of an utterance, and thereby conclude what type of error had been made, as well as preventing aborted errors such as *pung luse* → /… being counted as partial exchanges, thus reducing the error rate.

The paradigm did generate a large number of errors, as was predicted, with a total of 61.71% errorful utterances. Unfortunately, however, only 2.89% of all utterances were onset exchange errors, leaving little data for analysis. It must be concluded from this figure that the WOC task is not very efficient at generating onset exchange errors. There were, however, a much higher number of rime exchanges (9.14%). This pattern suggests that it is possible that the processing of reversing the pair was causing participants to pay particular attention to the correct placement of the onsets of the word, whilst the rimes were being more affected by the order confusion generated by the WOC task.

Rime exchanges could however also possibly provide ground for examining patterns such as the phonological similarity effect, by controlling for similarity in the coda. However, rime exchanges are not common in everyday speech (Motley, 1973). Whilst this does not mean that patterns of rime exchange errors would be of no value in investigating the underlying architecture and mechanism of the language production system, it would possibly be preferable to use more naturally occurring errors, such as onset exchanges.

Despite these issues, the WOC task does still appear to be the best choice of onset exchange error elicitation technique of those available today, as it avoids the articulatory and phonological biases present in other paradigms and is therefore more ecologically valid. Nevertheless, research would progress a lot faster if a new methodology could be
designed with none of the biases of the SLIP task, but a higher onset exchange error rate than that seen when using the WOC task.

**Extensions to the research**

Clearly, the most important extension to this research would be to perform a similar experiment, but with higher power, in order to truly determine how the lexical bias and phonological similarity effects interact. This could involve simply testing more people, or ideally, the use of a better error elicitation paradigm.

The work would also be well informed by carrying out a computational simulation of the feedback model, in order to investigate the exact form and strength of the interaction that this model would predict, as modulated by variables such as speaking deadlines. The usefulness of such modelling is underlined by the fact that Nooteboom (2005) had previously claimed that a feedback account would predict that lexical bias decreases as phonological similarity increases. Whilst this appears to be the result of a simple logical error, it underlines the fact that a pen and paper consideration of such complex models is sometimes not sufficient. Furthermore, such modelling could be combined with a more in depth examination of the proposals of Hartsuiker et al. (2005), examining how monitoring and feedback could be integrated into one model. Whilst their proposal is certainly interesting, it currently suffers from the plight of many monitor models, in that it is slightly underspecified. Solidifying their suggestions could generate interesting predictions suitable for further empirical examination.

An interesting issue which arose during examination of the data from our experiment was the role of orthography in speech error elicitation. Assessing the pattern of errors by items, it is noticeable that within the category of high phonological similarity, the items which are particularly conducive to producing errors also have the same first letter (for example, the pairs *tumb* thack and *sing* sull).

Damian and Bowers (2003) recently showed that orthography affects speech production using the implicit priming paradigm (Meyer, 1990). In this experiment, participants were asked to memorise pairs of words, such as *coffee-kennel*, or *coffee-cushion*. During the test period of the experiment, participants were given the first of a pair of words, and were
asked to produce the second. Experimental blocks were varied such that in some blocks, all responses began with the same orthographic and phonological segment, whereas in others, either the initial orthography was not the same, or neither the initial orthography nor the initial phonology was the same. Previous experiments of this type have shown that faster response times are seen in blocks in which phonology is kept the same. The results of Damian and Bowers’ (2003) experiment implied that in fact, this so-called phonological priming is only present when the initial orthography of the word is also maintained.

On the basis of these results, Damian and Bowers (2003) claimed that bidirectional connections exist between the orthographic and phonological subsystems. They suggested that for word production to occur, the language system must settle into a steady state. For this to occur, feedback from the orthographic system must be congruent with activation in the phonological system.

This work is relevant to the current topic as it highlights the importance that factors such as orthography may play in speech error patterns. It also considers yet another type of interactivity which may be present in the language system. Although any future work would have to very carefully account for the role of perceptual processes in the result patterns uncovered (c.f. Dell, 1986; Hartsuiker et al., Experiment 1b), it would be interesting to extend this work to speech error elicitation.

**Conclusions**

The results uncovered in the current experiment again confirm the presence of lexical bias and phonological similarity effects in phonological speech errors. They do not however support Nooteboom’s (2005) adapted monitor account of lexical bias and phonological similarity, which in its current incarnation is ruled out by the phonological similarity effect observed in lexical outcome errors.

Nevertheless, it should be noted that the trend towards lexical bias increasing with phonological similarity was only marginally significant. Feedback clearly could not account for a lack of interaction between the lexical bias and phonological similarity effect. Yet, on closer examination, it appears that the monitor model cannot account for
the phonological similarity effect, and therefore also could not explain such a result pattern, without postulating feedback.

Thus, this again implies that feedback is present in the language production system, and that the marginally significant result of lexical bias increasing with phonological similarity seen in this experiment is indeed indicative of an underlying effect. It is clear that a higher powered error elicitation paradigm would aid us in producing somewhat clearer data. However, given the limitations of the monitor model described in the discussion, it appears reasonable to conclude that the results detailed in this dissertation add to the ever increasing pool of evidence for feedback in language production.
References

References


**PHONOLOGICAL SIMILARITY ANDLexical Bias**
References


ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/dictionaries/beep.tar.gz


Acknowledgements

I am greatly indebted to Martin Corley and Corey McMillan for the amount of time and energy they devoted to supervising this project, and the knowledge and experience I have gained as a result. I would also like to thank Phil Collard, Natasha Dare and Zeynep Ilkin for useful comments and suggestions offered throughout the course of this work.
## Appendix 1: Materials

<table>
<thead>
<tr>
<th>Phonological Similarity Band</th>
<th>Frisch Value</th>
<th>Lexical</th>
<th>Nonlexical</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Stimulus</strong></td>
<td><strong>Outcome</strong></td>
</tr>
<tr>
<td>Low</td>
<td>0.03</td>
<td>meft thesh</td>
<td>theft mesh</td>
</tr>
<tr>
<td>Low</td>
<td>0.03</td>
<td>fock knнд</td>
<td>knock fond</td>
</tr>
<tr>
<td>Low</td>
<td>0.03</td>
<td>mave shain</td>
<td>shave main</td>
</tr>
<tr>
<td>Low</td>
<td>0.03</td>
<td>mun sunch</td>
<td>sun munch</td>
</tr>
<tr>
<td>Low</td>
<td>0.03</td>
<td>sile mide</td>
<td>mile side</td>
</tr>
<tr>
<td>Low</td>
<td>0.04</td>
<td>runk hust</td>
<td>hunk rust</td>
</tr>
<tr>
<td>Low</td>
<td>0.04</td>
<td>lund fump</td>
<td>fund lump</td>
</tr>
<tr>
<td>Low</td>
<td>0.04</td>
<td>pake laiнт</td>
<td>lake paint</td>
</tr>
<tr>
<td>Low</td>
<td>0.04</td>
<td>lulse pung</td>
<td>pulse lung</td>
</tr>
<tr>
<td>Low</td>
<td>0.05</td>
<td>fesk detch</td>
<td>desk fetch</td>
</tr>
<tr>
<td>Low</td>
<td>0.05</td>
<td>cush lup</td>
<td>lush cup</td>
</tr>
<tr>
<td>Low</td>
<td>0.06</td>
<td>fam jat</td>
<td>jam fat</td>
</tr>
<tr>
<td>Low</td>
<td>0.06</td>
<td>wrop shong</td>
<td>shop wrong</td>
</tr>
<tr>
<td>Low</td>
<td>0.06</td>
<td>datch hamp</td>
<td>hatch damp</td>
</tr>
<tr>
<td>Low</td>
<td>0.06</td>
<td>nug thut</td>
<td>thug nut</td>
</tr>
<tr>
<td>Low</td>
<td>0.08</td>
<td>guzz fum</td>
<td>fuzz gum</td>
</tr>
<tr>
<td>Low</td>
<td>0.08</td>
<td>gox fot</td>
<td>fox got</td>
</tr>
<tr>
<td>Low</td>
<td>0.08</td>
<td>cass mamp</td>
<td>mass camp</td>
</tr>
<tr>
<td>Low</td>
<td>0.09</td>
<td>rell setch</td>
<td>sell retch</td>
</tr>
<tr>
<td>Low</td>
<td>0.09</td>
<td>sotch noft</td>
<td>notch soft</td>
</tr>
<tr>
<td>Medium</td>
<td>0.11</td>
<td>thact panпк</td>
<td>pact thank</td>
</tr>
<tr>
<td>Medium</td>
<td>0.12</td>
<td>nait wame</td>
<td>wait name</td>
</tr>
<tr>
<td>Medium</td>
<td>0.12</td>
<td>taze hape</td>
<td>haze tape</td>
</tr>
<tr>
<td>Medium</td>
<td>0.13</td>
<td>jand lajз</td>
<td>land jazz</td>
</tr>
<tr>
<td>Medium</td>
<td>0.13</td>
<td>lem geft</td>
<td>gem left</td>
</tr>
<tr>
<td>Medium</td>
<td>0.13</td>
<td>huck chush</td>
<td>chuck hush</td>
</tr>
<tr>
<td>Medium</td>
<td>0.14</td>
<td>pife wike</td>
<td>wife pike</td>
</tr>
<tr>
<td>Medium</td>
<td>0.14</td>
<td>dunch puck</td>
<td>punch dusk</td>
</tr>
<tr>
<td>Medium</td>
<td>0.14</td>
<td>fouch cound</td>
<td>cоuch found</td>
</tr>
<tr>
<td>Medium</td>
<td>0.15</td>
<td>houch powl</td>
<td>pouch howl</td>
</tr>
<tr>
<td>Medium</td>
<td>0.15</td>
<td>gatch mang</td>
<td>match gang</td>
</tr>
<tr>
<td>Medium</td>
<td>0.15</td>
<td>gоuth mouge</td>
<td>mouth gouge</td>
</tr>
<tr>
<td>Medium</td>
<td>0.15</td>
<td>heg pedге</td>
<td>peg hedge</td>
</tr>
<tr>
<td>Medium</td>
<td>0.16</td>
<td>dush rull</td>
<td>rush dull</td>
</tr>
<tr>
<td>Medium</td>
<td>0.16</td>
<td>vap gan</td>
<td>gap van</td>
</tr>
<tr>
<td>Medium</td>
<td>0.17</td>
<td>tolf gop</td>
<td>golf top</td>
</tr>
<tr>
<td>Medium</td>
<td>0.17</td>
<td>mength leit</td>
<td>length melt</td>
</tr>
<tr>
<td>Medium</td>
<td>0.17</td>
<td>gax tas</td>
<td>tax gas</td>
</tr>
<tr>
<td>Medium</td>
<td>0.18</td>
<td>fout shouл</td>
<td>shout fouл</td>
</tr>
<tr>
<td>Medium</td>
<td>0.18</td>
<td>shoss tock</td>
<td>toss shock</td>
</tr>
<tr>
<td>High</td>
<td>0.20</td>
<td>tumb thuck</td>
<td>thumb tuck</td>
</tr>
<tr>
<td>High</td>
<td>0.22</td>
<td>choll dop</td>
<td>doll chop</td>
</tr>
<tr>
<td>High</td>
<td>0.23</td>
<td>hine shive</td>
<td>shine hive</td>
</tr>
<tr>
<td>High</td>
<td>0.23</td>
<td>sog holve</td>
<td>hog solve</td>
</tr>
<tr>
<td>High</td>
<td>0.23</td>
<td>celp hent</td>
<td>help cent</td>
</tr>
<tr>
<td>High</td>
<td>0.24</td>
<td>nunk jumb</td>
<td>junk numb</td>
</tr>
<tr>
<td>High</td>
<td>0.24</td>
<td>zell sheet</td>
<td>shell zest</td>
</tr>
<tr>
<td>High</td>
<td>0.25</td>
<td>choin coice</td>
<td>coin choice</td>
</tr>
<tr>
<td>High</td>
<td>0.25</td>
<td>kest chept</td>
<td>chest kept</td>
</tr>
<tr>
<td>High</td>
<td>0.27</td>
<td>samp chand</td>
<td>champ sand</td>
</tr>
<tr>
<td>High</td>
<td>0.27</td>
<td>gump juft</td>
<td>jump guft</td>
</tr>
<tr>
<td>High</td>
<td>0.27</td>
<td>jame gade</td>
<td>game jade</td>
</tr>
<tr>
<td>High</td>
<td>0.38</td>
<td>vell fent</td>
<td>fell vent</td>
</tr>
<tr>
<td>High</td>
<td>0.38</td>
<td>dudge nump</td>
<td>nudge dump</td>
</tr>
<tr>
<td>High</td>
<td>0.41</td>
<td>chelf sheck</td>
<td>shelf check</td>
</tr>
<tr>
<td>High</td>
<td>0.41</td>
<td>chall shap</td>
<td>shall chap</td>
</tr>
<tr>
<td>High</td>
<td>0.56</td>
<td>rounge lound</td>
<td>lounge round</td>
</tr>
<tr>
<td>High</td>
<td>0.56</td>
<td>lusk runch</td>
<td>rusk lunch</td>
</tr>
<tr>
<td>High</td>
<td>0.56</td>
<td>lut runge</td>
<td>rut lunge</td>
</tr>
<tr>
<td>High</td>
<td>0.58</td>
<td>shunk sut</td>
<td>sunk shut</td>
</tr>
</tbody>
</table>