Lexical selection and verbal self-monitoring: Effects of lexicality, context, and time pressure in picture-word interference

Elisah Dhooge & Robert J. Hartsuiker
Department of Experimental Psychology, Ghent University

Running head: lexical selection and monitoring

Address for correspondence:
Elisah D'Hooge
Department of Experimental Psychology
Ghent University
Henri Dunantlaan 2
9000 Ghent
Belgium
e-mail: Elisah.Dhooge@ugent.be

Abstract

Current views of lexical selection in language production differ in whether they assume lexical selection by competition or not. To account for recent data with the picture-word interference (PWI) task, both views need to be supplemented with assumptions about the control processes that block distractor naming. In this paper, we propose that such control is achieved by the verbal self-monitor. If monitoring is involved in the PWI task, performance in this task should be affected by variables that influence monitoring such as lexicality, lexicality of context, and time pressure. Indeed, pictures were named more quickly when the distractor was a pseudoword than a word (Experiment 1), which reversed in a context of pseudoword items (Experiment 3). Additionally, under time pressure, participants frequently named the distractor instead of the picture, suggesting that the monitor failed to exclude the distractor response. Such errors occurred more often with word than pseudoword distractors (Experiment 2); however, the effect flipped around in a pseudoword context (Experiment 4). Our findings argue for the role of the monitoring system in lexical selection. Implications for competitive and non-competitive models are discussed.

Introduction

A key process of language production is lexical access, the retrieval of lexical units containing syntactic and semantic information from the mental lexicon. There is consensus in the literature that due to a spreading activation mechanism at the conceptual level, not only the target unit but also categorically related units get activated (e.g., Caramazza, 1997; Dell, 1986; Levelt, Roelofs, & Meyer, 1999). Because of the multiple units activated at the lexical level, the speech production system needs a way to select the correct unit. The precise mechanism by which this proceeds is still debated. Generally, views of lexical selection differ in whether they assume lexical selection by competition or not. According to competitive models, the time it takes to select the correct lexical unit depends on the activation of all active lemmas (Bloem & LaHeij, 2003; Bloem, van de Boogaard, & LaHeij, 2004; Caramazza & Costa, 2000; Levelt, 1989; Levelt et al. 1999; Roelofs, 1992, 1993, 1997, 2001, 2003; Starreveld & LaHeij, 1995; Vigliocco, Lauer, Damian, & Levelt, 2002; Vigliocco, Vinson, Lewis, & Garrett, 2004). That is, the more lexical units are activated and the higher their activation is, the more difficult it will be and the more time it will take to select the correct unit. On the contrary, according to non-competitive models, lexical selection is independent of the activation of the non-target units (e.g., Caramazza, 1997; Dell, 1986). Only the activation of the target determines its selection: the higher its activation, the easier selection will be. Importantly, several empirical phenomena (see below) cannot be explained by the core principles of either competitive or non-competitive accounts; both types of account need to make further assumptions about the control processes that allow speakers to prevent interference. In this paper, we argue that such control is achieved by the verbal self-monitoring system, the system with which we inspect our own speech and initiate interruption and correction when necessary (Hartsuiker & Kolk, 2001; Levelt, 1989; Postma, 2000). If this monitoring proposal is correct, an account of key empirical data does not need to assume lexical selection by competition.

Competitive models were initially established and supported by the semantic interference effect in the picture-word interference (PWI) task (e.g., Schriefers, Meyer, & Levelt, 1990). In this task, participants see a picture with a superimposed distractor. They have to name the picture and ignore the distractor. When picture and distractor are semantically related (e.g., RAT-mouse), picture naming latencies will be prolonged compared to when they are unrelated (e.g., RAT – chair). Competitive models can account for this effect because a semantically related distractor will receive activation not only from the distractor, but also from the conceptual level due to the spreading activation. Unrelated distractors do not receive activation from this second source; hence, their activation will be lower. As a consequence, a semantically related distractor will be a more potent competitor, leading to longer naming latencies.

However, the empirical generalization that semantic relatedness always leads to interference has been challenged by the findings of several studies. In many conditions, semantically related distractors facilitate picture naming instead of leading to interference. For example, when the distractor and picture differ in grammatical class, semantic facilitation is found (Mahon, Costa, Peterson, Vargas, & Caramazza, 2007). That is, the verb ‘drive’ will speed up naming latencies of the picture of a car compared to the verb ‘sleep’. Another example is when manipulating the semantic distance between picture and distractor (Mahon et al., 2007; but see Vigliocco, Vinson, & Siri, 2008). Here, semantically close items (e.g., ZEBRA – horse) lead to shorter naming latencies compared to semantically further distractors (e.g., ZEBRA – cat). Other types of semantic relationships can also lead to facilitation. Abdel Rahman and Melinger (2007, see also Alario, Segui, & Ferrand, 2000) paired pictures with associatively related and unrelated distractors, and Costa, Alario and Caramazza (2005) used part-whole relationships. Both experiments showed shorter naming latencies with related compared to unrelated items.

Without making further assumptions, neither competitive models nor non-competitive models can account for this pattern of semantic interference with category coordinate distractors and semantic facilitation with other types of semantically related distractors. Given the consensus opinion that there is spreading of activation at the conceptual level, non-competitive models would in principle predict facilitation whereas competitive models would in principle predict interference (because conceptual priming increases competition). It is therefore necessary to postulate further mechanisms involved in lexical selection besides the core principle of competitive (or non-competitive) lexical selection. Similarly, another finding that cannot be explained by the core principles of either type of account is the distractor frequency effect. For example, Miozzo and Caramazza (2003; Dhooge & Hartsuiker, 2010; see also Burt, 2002) manipulated the competitive power of a distractor by using low- and high-frequency distractors. Similarly, Catling, Dent, Johnston and Balding, 2010, manipulated the age of acquisition of the distractor word. According to competitive models, high frequency and early acquired distractors should be stronger competitors (assuming their lexical representations have higher resting level activation) and so should lead to more interference. Non-competitive models would assume no influence of distractor frequency or age of acquisition. However, in contrast to both predictions low-frequency distractors interfered more than high-frequency ones (i.e., the distractor frequency effect). Likewise, late-acquired distractors produced more interference compared to early-acquired ones.

Together, these studies show that theories of lexical selection need to be extended with additional, specific assumptions. For instance, competitive models can be extended with a threshold that determines whether a lexical candidate enters the competition or not (e.g., Roelofs, Piai & Schriefers, 2011). Depending on the specific strength of the threshold, a lexical selection-by-competition account is then compatible with both semantic interference and semantic facilitation (and with a null effect, cf. Abdel Rahman & Melinger, 2009). Similarly, non-competitive models can be extended with a response exclusion mechanism that needs to exclude the response to the distractor from a speech output buffer before the response to the picture can be given (e.g., Miozzo & Caramazza, 2003). A further assumption is that this response exclusion process uses semantic information, which is quickly distilled from the picture, to determine that the buffered word is the name of the wrong stimulus. With such assumptions in place, a non-competitive account can also account for both semantic facilitation and semantic interference. A problem with such assumptions however is their lack of independent motivation and the lack of new predictions that follow from incorporating such assumptions. In this paper, we will claim that previous work has not considered the possible role of a device that is already in place in normal speech production, namely the verbal self-monitor. That is, the speech production literature has considered the possible effects monitoring has on the patterns of speech errors (e.g., Hartsuiker, 2006; Oppenheim & Dell, 2010) or on the patterns of disfluencies (e.g., Brocklehurst & Corley, 2011; Postma & Kolk, 1993) but so far has had little attention for effects of monitoring on the time-course of lexical selection processes. Below, we will first discuss monitoring mechanisms more in detail, highlighting three characteristics of monitoring that will allow us to make testable predictions. Next, we will explain the (tentative) role of the monitor in lexical selection and the PWI task, ending with an outline of the experiments.

Speech Monitoring

Speech monitoring is the on-line process of checking one's own speech for errors and other problems. This mechanism has been related to the more general action monitoring mechanism (e.g., Nozari, Dell, & Schwartz, 2011; see Ye & Zhou 2009, for a review). Importantly, it is often assumed that monitoring can intercept errors so that they can be corrected covertly, leaving resulting speech fluent. Although the literature is divided on the precise error detection mechanisms (e.g., Huettig & Hartsuiker, 2010; Ozdemir, Roelofs, & Levelt, 2007) there is a consensus that people monitor not only their overt speech but also a representation of pre-articulatory speech. First, some speech errors are intercepted too fast (i.e., within 150ms) to be based on overt speech (e.g., Blackmer & Mitton, 1991). Second, Oppenheim and Dell (2008) asked participants to recite tongue twisters, either as overt or covert speech. Importantly, errors were detected in both conditions. Third, errors can be detected when overt speech cannot be monitored as shown by studies using noise masking (Lackner & Tuller, 1979; Postma & Kolk, 1992; Postma & Noordanus, 1996). Fourth, a study by Severens, Janssen, Kühn, Brass and Hartsuiker (in press) showed that word pairs that could have yielded an embarrassing taboo error but were named correctly (e.g., 'tool kits' that could have been read as 'cool tits') yielded an effect on ERPs. These data have been taken as evidence that the taboo sequence was formulated internally but subsequently was detected and corrected by the verbal self-monitor.

Previous work on self-monitoring suggested three features of the monitoring system which are relevant for our purposes, namely a sensitivity to lexicality, to context, and to time pressure. First, the monitor checks speech on a number of factors, such as social acceptability, syntax, morphology, etc. (for a detailed overview, see Levelt, 1989). Importantly, the monitor also is assumed to use lexical status (or lexicality, i.e., 'is it a word') as a criterion (but see Nozari & Dell, 2009). In the context of normal speech production, this is a functional criterion, as violations against it would result in nonsense speech. The use of this criterion is supported by the lexical bias effect both in corpus studies (e.g., Hartsuiker, Anton-Mendez, Roelstraete, & Costa, 2006; Nooteboom, 2005b) and in the SLIP task (Spoonerisms of Laboratory Induced Dispositions; e.g., Baars, Motley, & MacKay, 1975; Dell, 1986; Hartsuiker et al., 2006; Humphreys, 2002; Nooteboom, 2005a; Nooteboom & Quené, 2008). For example, in the SLIP task, word pairs that have to be read aloud are preceded by word pairs that bias them into making errors, like exchanging the first phonemes of each word (tool carts -> cool tarts). The lexical bias effect then refers to the observation that errors resulting in existing words will occur more than errors resulting in pseudowords. This lexical bias effect follows naturally from a lexicality criterion (i.e., 'is it a word'). When adhering to a lexicality criterion, word errors are more likely to pass the criterion than pseudowords, leading to more word errors. Put differently, the monitor does not catch all errors with equal probability but exhibits a monitoring bias.

A second feature is the context-dependent nature of the monitor. This context-specificity was first shown in Baars et al.'s (1975) Experiment 2 where target and biasing word pairs were always pseudowords. These authors created a mixed context by inserting real word fillers and a pseudoword (i.e., non-lexical) context by presenting only pseudoword fillers. A lexical bias effect was only found in the mixed context. According to Baars et al.'s interpretation, the monitor would use the lexicality criterion in the mixed context, but it would refrain from doing this in a pseudoword context. Converging evidence for a context-sensitive monitor comes from Hartsuiker, Corley and Martensen (2005). These authors went one step further and argued that a truly adaptive monitor should not only give up the lexicality criterion in the pseudoword context but even adapt the reverse "non-lexicality" criterion (i.e., the criterion: is this a pseudoword?). In a better controlled experiment using the same basic paradigm as Baars et al. (1975), they found that in the non-lexical context, the number of word errors was reduced compared to the mixed context whereas the number of non-lexical errors remained unchanged. They explained the form of their lexicality by context interaction as a result of the interplay between an adaptive monitor and a feedback mechanism (also see Nooteboom & Quené, 2008; Oppenheim & Dell, 2010). That is, in a pseudoword context, the monitor should quickly pick up the nature of the context and set its criteria accordingly. However, in the mixed context, neither the lexicality nor the reverse criterion is useful, as subjects encounter both words and pseudowords. Thus, the data in that condition reflect the amount of word and pseudoword errors without monitoring, namely more lexical than non-lexical errors. Hartsuiker et al. accounted for this latter finding by assuming a feedback mechanism. Combined with the non-lexicality criterion in the pseudoword condition, this will establish the pattern of suppression of lexical errors in the non-lexical context.

Finally, the monitor needs time to perform its task well (Hartsuiker, 2006, see also Dell, 1986). In one study that has directly investigated this issue, participants had to describe a network consisting of pictures connected with straight or bended lines (Oomen & Postma, 2002, also see Jou & Harris, 1992). They did this either with or without generating random sequences of finger taps, a task which requires attention and is disruptive to the operation of the central executive (cf. Baddley, 1996; Baddeley, Emslie, Kolodny, & Duncan, 1998). In the divided attention condition, there were more errors and fewer self repairs than in the single task condition. These results show that when other tasks take up processing time, the functionality of the monitor decreases. A second experiment of interest considered event-related potentials, and more specifically, the error-related negativity (i.e., ERN, Ganuskchak & Schiller, 2006). This component has a fronto-central scalp distribution and peaks approximately 80 ms after an overt incorrect response; it is often seen as related to error monitoring processes (e.g., Riès, Janssen, Dufau, Alario, & Burle, 2011). Participants monitored their internal speech in a phoneme monitoring task. The amplitude of the ERN decreased under time pressure, suggesting less monitoring activity under these circumstances (cf. Ghering, Goss, Coles, Meyer, & Donchin, 1993). This was corroborated by the fact that phoneme monitoring was also more error prone under time pressure. Third, Dell (1986) found that the lexical bias effect in the SLIP task disappeared under strong time pressure. One interpretation is that the time pressure prevented participants from using a criterion of lexicality.

In sum, the monitor is seen as a device that uses lexicality as a criterion, is context dependent, and needs time to perform its job well. In the experiments presented below we will test whether these three properties of the monitor affect participants' performance in the picture-word interference task.

Lexical selection and monitoring

As the goal of the monitor is to keep speech error-free, one of its tasks is to intercept and correct lexical errors. In other words, we assume that the monitor can intercept errors before they are made, leading to the absence of overt errors. Therefore, it makes a vital contribution to smooth lexical selection processes. We will examine the role of the monitor by means of the PWI task. By introducing the distractor, the PWI task is a useful task to study speech production and monitoring. After all, in daily life, people are continuously bombarded with irrelevant stimuli, that could potentially interfere with speech production. This situation is excellently mirrored in the PWI task. Specifically, in the PWI task, participants will activate two lexical units: the one of the picture and the one of the distractor, increasing the likelihood of erroneous lexical selection. Then, the task of the monitor will be to prevent errors by detecting the (potential) error and correct it before it is spoken out loud.

Note that the monitor has a natural account for several findings reported in the literature. In particular, several studies that manipulated distractor lexicality or time pressure in the PWI task are consistent with the hypothesis that the monitor has a role in the PWI task. First, early PWI studies demonstrated that a pseudoword interferes less than a word (e.g., Klein, 1964; Lupker 1979; Rosinski, Golinkoff, & Kukish, 1975), although it is unclear from those papers whether the distractors were matched for the same sort of variables that are typically controlled in more recent papers. This finding is consistent with the lexicality criterion of the monitor. Because a pseudoword is detected as inconsistent with the task at hand more easily, it can be excluded as a response earlier. A second line of research focused on effects of time pressure. Speeding up the PWI task (i.e., name the picture as fast as possible without worrying about errors) should hamper functioning of the monitor (e.g., Oomen & Postma, 2002). In particular, whereas normally the monitor has enough time to check responses, this should not be the case under time pressure. Also, as the distractor is presumably the word with the highest lexical activation, naming of the distractor is the most likely candidate for a speech error. Thus, when encouraging subjects to answer very quickly, one should observe distractor naming errors. Indeed, Starreveld and LaHeij (1999) found distractor naming errors (i.e., naming the distractor instead of the picture) under these time pressure conditions. In line with a monitoring account, these distractor naming errors differed over distractor conditions. There were more distractor naming instances when the distractor was relatively similar (i.e., semantically or phonologically related) to the target (Starreveld & La Heij, 1999).

The idea that the monitor is involved in the PWI task bears strong resemblance to the response exclusion hypothesis (e.g., Dhooge & Hartsuiker, 2010; 2011a-b; Finkbeiner & Caramazza, 2006; Janssen, Schirm, Mahon, & Caramazza, 2008; Mahon et al., 2007; Miozzo & Caramazza, 2003). The response exclusion hypothesis assumes that speakers first spuriously form a response to the distractor word which then needs to be discarded from a response buffer, as words have a privileged access to the articulators. Because of the need to purge the distractor, naming latencies depend on two factors: the time the distractor response enters the buffer and the time it takes to remove its response from the buffer. Response latencies become faster as the distractor response enters the buffer faster and if it can be removed from the buffer faster. That the first factor plays a role is supported by the distractor frequency effect (e.g., Dhooge & Hartsuiker, 2010; Miozzo & Caramazza, 2003; see also Burt, 2002). Because responses to low-frequency words are available later, for example due to a lower resting activation level (Dell, 1986), they will enter the buffer later compared to high-frequency distractors. The semantic interference effect can be explained by the second factor. Upon picture processing, the control mechanism is assumed to have semantic information at its disposal, which can help to decide whether the distractor response is correct or not. In the case of a semantically related word, this information will not be useful because it does not discriminate between target and distractor. Therefore, it will take more time to reject the response from the buffer. Summarizing, the response exclusion hypothesis assumes that during the PWI task, a control process operates over a response buffer and checks responses for accuracy. This way, effects in the PWI task can (sometimes) be situated at a post-lexical level without the need to invoke a competitive lexical selection mechanism. Note that this control process can be easily equated to the verbal self-monitor, as has been previously suggested (Dhooge & Hartsuiker, 2010).

Outline of the Experiments

The goal of the experiments was to find evidence for a role of self-monitoring in safeguarding the accuracy of lexical selection. More in detail, if we can show that performance in the PWI-task is sensitive to variables that affect self-monitoring, it is reasonable to assume that the self-monitor is the system that takes care of intercepting responses formed to distracting information, and that it is this system with which any theory (competitive or not) should be extended. To this extent, we present four experiments. To show and replicate the sensitivity to lexical status, all experiments employ a PWI task with word and pseudoword distractors. Words should then interfere more than pseudowords. Additionally, we seek evidence for the monitoring hypothesis by checking if there are any effects of time pressure by introducing a speeded version of the task. We expect that under such conditions participants start naming the distractor, especially if that distractor is relatively difficult to exclude by the monitor (Experiment 2 and Experiment 4). Specifically, there should be more word than pseudoword distractor naming errors, because pseudoword distractor naming errors should be easier to intercept. Third, these results should vary with context. In a context of pseudowords, a 'non-lexicality criterion' is more functional than a lexicality criterion. If the monitor indeed adjusts its criterion in this way, word distractor naming errors should now be more likely to be intercepted. This should result in more distractor naming errors and longer naming latencies with pseudoword distractors (Experiment 3 and Experiment 4).

Experiment 1: Distractor Lexical Status Effect in Response Times

In Experiment 1, the goal was to replicate the finding that in a PWI task words interfere more than pseudowords when we control for variables that were not generally controlled in the nineteen seventies (e.g., Lupker, 1979; Rosinski et al., 1977).

Method

Subjects. Twenty undergraduate students of Ghent University took part in the experiment and received €5 for their participation. All subjects reported normal or corrected to normal vision and were native speakers of Dutch. They were all naïve to the purpose of the experiment. None of them participated in any of the further experiments.

Design. Naming latency was considered as the dependent variable. Lexical status of the distractor was the within-subjects and within-item independent variable. It included two levels: word or pseudoword.

Materials. Sixty-two black and white drawings of concrete objects and items were selected from the Severens, Van Lommel, Ratinckx and Hartsuiker (2005) database. This database contains timed picture naming norms for Dutch for 590 pictures together with a number of variables know to influence naming latencies such as frequency stemming from the Celex database (Baayen, Piepenbrock, & Van Rijn, 1993).

Each picture was paired with a word and a pseudoword. All word distractors were nouns to avoid any influence of grammatical class. Pseudoword distractors were generated using WordGen (Duyck, Desmet, Verbeke, & Brysbaert, 2004). Distractors and pictures used in the practice phase were never used in the experimental phase. Both the distractors used in the practice phase and the distractors used in the experimental phase were semantically and phonologically unrelated to the picture with which they were paired. To create the practice stimuli, 12 pictures were paired up with a word and pseudoword. Fifty pictures each paired up with a word and a pseudoword formed the 100 experimental stimuli. Details on the selected stimuli can be found in Table 1. The stimuli themselves can be found in the Appendix. Word and pseudoword pairs always had an equal number of letters. Furthermore, word and pseudoword distractors were matched on bigram frequency, t(49) = .82, p = .42; number of neighbors, t(49) = .55, p = .58; number of syllables, t(49) = -1.55, p = .13; and number of phonemes, t(49) = -1.00, p = .32.

Insert Table 1 about here

A plus sign served as fixation point. Stimuli were presented centrally on screen. Distractors were presented centrally as well, so that the center of each word was at the center of each picture. All stimuli were presented centrally on a 17 in. monitor with a 60Hz refresh rate placed at a distance of 60 cm in front of the subject. Distractors were presented centrally in black capital letters in a Times New Roman font 26 points. Pictures were 300 x 300 pixels large and appeared centrally on screen. Stimulus delivery and millisecond accurate response registration was achieved by means of the Tscope package (Stevens, Lammertyn, Verbruggen, & Vandierendonck, 2006) run on a Pentium 4PC. Oral responses were collected through a NEVK voice key (Duyck, Anseel, Szmalec, Mestdagh, Tavernier, & Hartsuiker, 2008) connected to the parallel port.

Procedure. The experiment was run in a sound attenuated, dimly lit room and consisted of two phases. In a first phase, participants were familiarized with the names of the pictures. On-screen written instructions that were verbally clarified informed the participant that they were to name the picture when the name appeared below it. They were told to study the picture attentively as they had to use the correct name in the next phase. A trial started with a fixation cross for 700 ms. Next, the picture was presented for 1000 ms after which the name was presented below. After the subject named the picture, the experimenter typed in whether the response was correct or not. The next trial was initiated 1000 ms afterwards.

After a self-paced break, the second phase started. It consisted of a practice phase and an experimental phase. Each trial started with a fixation cross for 1000 ms followed by a blank screen for 500 ms. Then the picture and its distractor appeared for 350 ms. After the subject responded, the experimenter typed in whether the response was correct or not. This initiated the next trial after 500 ms. Picture and distractor presentation was relatively short, to speed up responses. This was especially important in Experiments 2 and 4, but we opted to keep the stimulus timing constant across experiments. Subjects were asked to name the picture as accurately and fast as possible whilst ignoring the distractor. The practice block consisted of 24 trials. In this phase, each picture was presented twice: once with its word distractor and once with its pseudoword distractor. In the experimental phase, participants ran through two blocks of trials. In each block, the picture was shown once with its pseudoword and once with its word distractor. This resulted in 100 trials per block, leading to 200 trials in total. Trials were randomized per block with the restriction that (a) a picture could not be repeated before all other pictures were named once, (b) stimuli in the same lexical condition could not appear on more than three consecutive trials and (c) per set of 10 trials there had to be five pictures paired with a pseudoword and five with a word. Furthermore, whether a certain picture would appear with its word or pseudoword distractor first was counterbalanced across subjects. Subjects could take a self-paced break between the practice and the experimental phase and between blocks.

In total, the experiment lasted approximately 30 minutes.

Results and discussion

Errors and outliers were removed from the dataset. Errors included malfunctioning of the voice key and wrong naming of the picture. However, no naming errors were made. Outliers were defined as all reaction times exceeding the participant's mean by three standard deviations, and all reaction times smaller than 300 ms. Errors and outliers accounted for 4.23% and 1.93% of the data. Error rates are typically low in this paradigm and were not further analyzed. These criteria were also used in Experiment 3. Means and standard deviations can be found in Table 2**.**

Insert Table 2 about here

Linear mixed models were fitted to the data, as implemented in the *lme4* library (Bates, 2007) in R (R Development Core Team, 2009, for comparison, the results of repeated-measures ANOVAs are presented in Appendix 2). In this and all other experiments, the null model included a random intercept for participants and items. A χ-test showed that including a random slope did not increase the fit of the model, χ^2 (2) = .16, p = .92, so only random intercepts were included. Importantly, including lexical status as the independent variable improved the fit of the model significantly, χ^2 (1) = 20.62, p < .001. A t-test showed that the difference between conditions was significant, ß = 21.83, SE = 4.80, t = 4.55, p < .001. Naming latencies were 22ms longer with word than pseudoword distractors.

In line with our hypothesis and with previous literature, words interfered more than pseudowords. This finding is consistent with the hypothesis that the monitor deals with the distractor words. That is, as pseudoword errors are detected faster they can be removed faster, leading to faster naming latencies. The next experiment tests whether monitoring can sometimes fail given strong enough time pressure, so that incorrect responses that are relatively difficulty to exclude (word distractor naming errors) are produced more often than incorrect responses that are more easy to catch (pseudoword distractor naming errors).

Experiment 2: Distractor Lexical Status Effect in Distractor Intrusions

Experiment 2 tested whether the mechanism that deals with the distractor in the PWI task is sensitive to time pressure. In a speeded version of the PWI task (cf. Starreveld & LaHeij, 1999), subjects were asked to name the picture as quickly as possible, without worrying about errors (i.e., naming the distractor instead of the picture). We expected this manipulation to hamper the functioning of the monitor, resulting in more distractor naming errors. In particular, more distractor naming errors were expected with word compared to pseudoword distractors.

Method

Subjects. Twenty further undergraduate students of Ghent University took part in the experiment and received €5 for their participation. All subjects reported normal or corrected-to-normal vision and were naïve to the purpose of the experiment. They were all native speakers of Dutch.

Design, Materials, Procedure. The design, materials, and procedure were identical to Experiment 1 except for the instructions for the practice and the experimental phase. To elicit errors on picture naming trials, similar to Starreveld and La Heij (1999), subjects were told that they had to name the pictures as fast as possible, without worrying about errors. Furthermore, they were told that if they were not making any mistakes, they were responding too slowly. Additionally, between blocks and between the practice and experimental phase, participants were reminded of these instructions. They were told that if they made an error, they should not correct themselves. This was done to keep the pace of the experiment high. Also, to prevent hesitations and small vocal sounds, participants ran through the practice block until they did not make any of them anymore. Additionally, they were told not to withhold themselves in case they felt they were about to make an error. This was done to prevent subjects from silently making an error and then correcting it.

Results and discussion

Fifty-seven out of all 4000 responses (1.43%) were scored as other than correct naming of the picture or the distractor. All of these other responses were verbal hesitations which were then followed by a picture or distractor naming response. Excluding these responses from the dataset did not make any differences to the results and so they were retained in the dataset. The percentage of distractor naming can be found in Table 3**.** Responses were fitted using a mixed logit model that predicts the logit transformed likelihood of a picture naming response as implemented in the *lme4* library (Bates, 2007) in R (R Development Core Team, 2009). We included a random intercept for subjects and items.

A χ^2-test was used to determine whether the inclusion of a random slope significantly increased the fit of the model, which it did, χ^2 (2) = 17.74, $p < .001$. As in Experiment 1, including the lexical status significantly improved the null model, χ^2 (1) = 42.23, $p < .0001$. A Wald z-test showed that this difference between conditions was significant, ß = 2.02, SE = 0.17, z = 11.81, $p < .001$. Thus, participants were more likely to name the distractor when it was a word.

Insert Table 3 about here

The findings support our predictions. The distribution of distractor naming errors points to the involvement of the monitor. According to the monitoring hypothesis, time pressure should reduce the likelihood that the monitor blocks out distractor naming errors, leading to the production of overt errors. Most importantly, the fact that we found more word than pseudoword distractor naming errors supports a monitoring account. As pseudoword (distractor naming) errors are easier to detect and rule out than word (distractor naming) errors, more word distractor naming errors should be made. Thus, the present distractor lexicality effect can be considered as an equivalent in the PWI-task to the lexical bias effect in the SLIP task (e.g., Baars et al., 1975).

Note that a shortcoming of this design is that due to the limitations of the recording apparatus used in this experiment (cfand also in. Experiment 4), no naming latencies could not be recorded. Yet, given that the participants did make errors (in contrast to they did not make in Experiment 1), we believe there is no reason to question the effectiveness of the instructions in inducing time pressure.

The monitor has been argued to be a flexible device that is adaptive to the speaking situation at hand (Levelt, 1989) and indeed the lexical bias effect in the SLIP task is modulated by context (Baars et al., 1975; Hartsuiker et al., 2005). Therefore, the next two experiments investigated whether the effect of lexical status of the distractor on response times and distractor intrusions shown in Experiments 1 and 2 are also context-dependent. If so, this would be strong support for our monitoring account.

Experiment 3: Distractor Lexical Status and Context Effects in Response Times

In Experiment 3, the goal is to test whether the effect of lexical status of the distractor on response times shown in Experiment 1 is subject to context effects. If the monitor is involved in the PWI task, it should behave adaptively and adjust its monitoring criteria according to the context (e.g., Hartsuiker et al., 2005). Therefore, we included either word or pseudoword fillers that had to be named. In this way, we created a condition in which only words or mostly pseudowords (75% of trials) have to be pronounced. In a word context, the most adaptive criterion is lexicality so that we should replicate the results from Experiment 1. However, in the pseudoword context, mostly pseudowords have to be named. Then, either a non-lexicality criterion (Hartsuiker et al., 2005) or no monitoring criterion at all would be optimal. Therefore, in the pseudoword context we expect either more interference from pseudowords than words or equal interference.

The manipulation of context did require a change to the design of Experiments 1 and 2. In Experiments 3 and 4, lexical status of the distractor was manipulated between-blocks whereas it was manipulated within-blocks in Experiments 1 and 2. This was done to avoid surprise effects. That is, when naming pictures in the context of word (pseudoword) fillers, a pseudoword (word) distractor is the odd-one-out. This rare item could create a surprise effect, which would result in longer naming latencies for pictures paired with a pseudoword in the word context and with a word in the pseudoword context. Keeping lexical status of the distractor constant in each block rules out that the results are influenced by surprisal.

Subjects. Thirty-six further undergraduate students of Ghent University took part in the experiment for partial course credit. All subjects reported normal or corrected-to-normal vision and were naïve to the purpose of the experiment. They were all native speakers of Dutch.

Design. Naming latency was the dependent variable. There were two independent variables. Lexical status of the distractor was varied within-subjects and within-items. Lexical status of the context was manipulated between-subjects and within-items. Both variables had two levels: word or pseudoword.

Materials. Materials were identical to Experiment 1 and Experiment 2 with two exceptions. First, the practice phase now consisted of 24 pictures paired with a word and pseudoword, giving rise to 48 trials. These were intermixed with 72 words or 72 pseudowords (i.e., fillers). Second, context (filler) stimuli were selected for the experimental phase as well. These stimuli consisted of 200 words and 200 pseudowords. All words were nouns, to avoid any effects of grammatical class. Pseudowords were generated using WordGen (Duyck et al., 2004). Details on the selected word and pseudoword fillers are presented in Table 1**.** Word and pseudoword fillers were pair-wisely matched on number of letters. Furthermore, there were no significant differences regarding the bigram frequency, t(199) = -.10, p = .92; the number of neighbors, t(199) = .75, p = .45; number of syllables, t(199) = .58, p = .57; or the number of phonemes, t(199) = .19, p = .85. Fillers were never used as distractors.

Procedure. The procedure was identical to that of Experiment 1 and Experiment 2, with the following exceptions. First, picture naming trials were intermixed with trials in which a single word or pseudoword (i.e., fillers) had to be read out loud. In the word context, fillers always consisted of real words whereas in the pseudoword context fillers were always pseudowords. Furthermore, lexical status of the distractor was manipulated between blocks. Thus, in each block, the distractor would always be either a word or a pseudoword. The order of blocks was counterbalanced across participants. A practice phase was now administered before each experimental block. Each practice phase now consisted of 96 trials and each experimental block consisted of 250 trials, leading to 500 trials in total. On-screen written instructions informed the participant that there were two types of trials: word or pseudoword trials (depending on the condition the subject was in) and picture naming trials. The subjects were instructed to read the word or pseudoword aloud when it appeared alone, and to name the picture whilst ignoring the distractor when a picture appeared. Trials were randomized per block with the additional restriction that (a) a picture could not be repeated before all other pictures were named once and (b) there had to be at least one and at most eight filler trials between two picture naming trials. This last restriction ensured that all picture naming trials were preceded by a filler trial, so that any effect of task switching would affect all conditions equally. Trial structure was identical to the two previous experiments with the exception that now either the filler item or the picture with its distractor appeared for 350 ms.

In total, the experiment lasted about 45 minutes.

Results and discussion

The same criteria for data-trimming were used as in Experiment 1. Only picture naming trials were analyzed. Errors and outliers accounted for 2.44% and 3.78% of the data. Means and standard deviations for the different conditions can be found in Table 2.

As in Experiment 1, we fitted linear mixed models to the data, as implemented in the *lme4* library (Bates, 2007) in R (R Development Core Team, 2009), with a random intercept for participants and items. A χ^2-test showed that including a random slope for participants significantly improved the fit of the model, $\chi^2(9) = 39.88$, $p < .001$, so this parameter was included as well. Importantly, a model comparison showed that the model that included the interaction term had to be preferred over a model with only main effects, $\chi^2(1) = 14.89$, $p < .001$, indicating that the effect of lexical status of the distractor varies between conditions.

To examine this interaction more in detail, we fitted the data of each condition separately. A model with random slope for participants significantly improved the fit of the model for words, χ^2 (2) = 27.40, $p < .0001$, and for pseudowords, χ^2 (2) = 12.06, $p < .01$. In line with the hypothesis, including lexical status as a predictor improved the fit as well, both for the word context, χ^2 (1) = 6.08, $p < .05$, and for the pseudoword context, χ^2 (1) = 9.04, $p < .0001$. These results indicate that word distractors interfered more than pseudoword distractors in a context of words (difference = 36ms) whereas the effect reversed in a context of pseudowords (difference = -42ms). This difference was significant in both the word context, ß = 35.58, SE = 13.23, t = 2.69[i], and the pseudoword context, ß = -41.66, SE = 8.32, t = -5.01, p < .001.

The data of Experiment 3 are clearly in line with a monitoring account. They extend the results of Experiment 1 by showing that the interference effects of words and pseudowords vary according to the context. That is, in a word context, subjects appear to use a lexicality criterion. This entails that a word distractor naming error is more difficult to detect, resulting in longer naming latencies compared to pseudowords. However, in a pseudoword context, subjects switch to a non-lexicality criterion. That is, they monitor according to the criterion "is it a pseudoword"? As a result, pseudoword distractor naming errors should be more difficult to detect, resulting in longer naming latencies compared to word distractors. Note that if subjects would have abandoned any lexicality criterion, we should have found equal interference of words and pseudowords. In a final experiment, we further tested the involvement of the monitor by considering distractor naming errors.

Experiment 4: Distractor Lexical Status and Context Effects in Distractor Intrusions

Experiment 4 tested for the effects of all three hypothesized features of the monitor at once. We therefore combined the methods of Experiment 2 and Experiment 3 to see whether the context dependability of the control mechanism can be extended to distractor naming errors as well. Thus, we expect that in a word context, we should replicate Experiment 2. As Experiment 3 found evidence that subjects use a non-lexicality criterion to monitor responses, we now expect more pseudoword than word distractor naming errors in a pseudoword context.

Method

Subjects. Thirty-six further undergraduate students of Ghent University took part in the experiment for partial course credit. All subjects reported normal or corrected to normal vision and were naïve to the purpose of the experiment. They were all native speakers of Dutch.

Design, Materials, Procedure. The design, materials, and procedure were identical to Experiment 3 except for the instructions, which were identical to those of Experiment 2 and thus emphasized speed over accuracy. We also included a break after each 50 trials.

Results and discussion

Responses were scored as picture or distractor naming responses. No other type of response was given. Percentage of distractor naming can be found in Table 3. Responses were fitted using a mixed logit model that predicts the logit transformed likelihood of a picture naming response as implemented in the *lme4* library (Bates, 2007) in R (R Development Core Team, 2009) with a random intercept for participants and items. Including a random slope for participants did not improve the fit of the model, χ^2 (9) = 4.39, p = .88. As in Experiment 3, a model with the main effects of lexical status of the distractor and lexical status of the context, and their interaction fitted the data better than a model without, χ^2 (1) = 61.79, p < .001.

To investigate this interaction more closely, we again fitted logit mixed models with a random intercept for participants and items for each condition separately. For both the word and pseudoword condition, including a random slope did not improve the fit of the model, χ^2 (2) = 0.05, p = .97, χ^2 (2) = .30, p = .86, respectively. In line with the hypothesis, lexical status made a significant contribution to the model both in the word condition and in the pseudoword condition, χ^2 (1) = 51.76, p <.001, χ^2 (1) = 19.93, p < .001, respectively. A Wald z-test showed that the difference was significant in both conditions as well, ß = -1.11, SE = .02, z = -6.87, p < .001, ß = .62, SE = .14, z = 4.44, p < .001, respectively.

This final experiment forms the most solid piece of evidence for the role of the monitor in the PWI task. It demonstrates the three properties associated with the verbal self-monitor. First, words interfere more than pseudowords. Second, this effect depends on the lexical status of the context: if more pseudowords need to be named, the effect reverses. Finally, these results have been obtained under time pressure, resulting in speech errors.

Note that in Experiment 3, on average, naming latencies increased compared to Experiment 1, suggesting that the task in Experiments 3-4 (involving task switching) is more difficult than that in Experiments 1-2. But this is not consistent with the error data: . ~~However, this pattern reversed for Experiment 2 and 4:~~ more errors were made in Experiment 2 compared to 4. A possibility is that this again shows involvement of the monitor. That is, the more difficult the task, the more stringent the monitor will be, resulting in fewer errors and increased naming latencies. Another question that may arise is whether the context effects of Experiment 3 and 4 ~~are not merely due~~should be attributed to a task switching or spill-over effect. That is, when the previous filler and the distractor share lexical status (e.g., both pseudoword), it might be more difficult to exclude the distractor. This account would then predict longer naming latencies and more distractor naming errors in Experiments 3 and 4 compared to Experiments 1 and 2, respectively. However, the data show that whereas this account is consistent with the naming latencies, it does not explain the higher average error rate in Experiment 2 compared to Experiment 4.

Finally, it could be argued that our instructions encouraged subjects to name the distractor. This, indeed, is an option that cannot be excluded. However, this would only mean that the amount of distractor naming errors is overestimated, but does not change the fact that the number of distractor naming errors differed across conditions (i.e., word vs. pseudoword

Opmerking [RH1]: Kun je eigenlijk niet zeggen – het ene is een RT patroon, het andere een foutpatroon.

Met opmaak: Markeren

distractor, and the reversal according to lexical status of the context). This is especially true given the fact that our instructions were the same in each condition and for each participant.

General discussion

Current theories of lexical selection do not assign a role to the verbal self-monitor, a control mechanism that checks our fluent speech for well-formedness and detects errors so they can be filtered out. Four experiments tested whether there is such a role, exploiting three hypothesized properties of the monitor. First, the monitor checks our speech according to the lexicality criterion (i.e., ‘is this a word’) in order to prevent nonsense speech. Second, the monitor needs time to check responses, making it sensitive to time manipulations. Third, the monitor is flexible, changing its criteria according to the context. That is, when the context consists of words, monitoring according to lexicality is the best way to go. However, when (mostly) pseudowords have to be produced, no lexicality criterion or even the reverse, a non-lexicality criterion is optimal. Thus, if the monitor is involved in lexical access, we should observe a sensitivity to these three properties in the picture-word interference task.

Our experiments indeed found evidence that all three monitoring properties affected performance in the PWI task. Experiment 1 and Experiment 3 (word context) showed that pseudoword distractors interfere less than word distractors in a picture naming task, which is in line with the use of a lexicality criterion. If pseudoword (distractor naming) errors are detected and removed faster, the pictures they accompany should be named faster. An effect of the second property, sensitivity to time pressure, was demonstrated in Experiment 2 and Experiment 4 (word context). In these experiments, participants performed a speeded version of the PWI task. The rationale behind this approach is that when the monitor has less time to perform its task, it will also perform it less accurately. However, if pseudoword errors can be detected easier, they should be affected less by this reduced monitoring than word errors. In line with this prediction, more distractor naming errors were made when the distractor was a word compared to a pseudoword. Finally, Experiments 3 and 4 showed that the control mechanism is sensitive to the context in which speech is produced. When naming pictures was intermixed with trials in which single words were named, the results mirrored those of Experiment 1 and 2 where only pictures had to be named. There again were longer naming latencies and more distractor naming errors with word compared to pseudoword distractors, which is in line with a lexicality-based monitoring criterion. However, when the context consisted of pseudowords, and thus mostly pseudowords had to be named, these effects reversed. There was more distractor naming with a pseudoword than with a word distractor and picture naming latencies were slower with pseudoword distractors as well. The results of the pseudoword context can be explained best by a non-lexicality criterion, so that now word distractor naming errors are detected faster. Together, these experiments form a test of the role of the monitor in lexical access. Previous research found some evidence compatible with the hypothesis that the monitor is involved in the PWI task, consistent with our findings. However, to the best of our knowledge, no study has yet addressed the issue of context specificity.

In sum, these findings indicate that the self-monitoring system helps the lexical access process stay on track, and so is a factor that should not be neglected in the interpretation of PWI data and related paradigms. Importantly, neither a non-competitive model nor a competitive model seems able to explain the complete data pattern without assuming a role for the monitor, even though they might be able to explain part of the findings. First, consider the effects on error rates (Experiment 2 and 4-word context). These effects can be incorporated in both types of models when starting from their core assumptions (i.e., competition or not). That is, given that the word distractor is highly activated, both a competitive and non-competitive models can incorporate the idea that by chance this lexical unit will be selected instead of the correct one. As a pseudoword does not have a lexical entry, these errors will occur less frequency. A competitive model can also explain the effects on naming latencies (Experiment 1 and 3-word context). After all, a word would be a viable competitor but a pseudoword would not, leading to longer naming latencies with words compared to pseudowords. A non-competitive model, on the other hand, cannot explain why a word distractor would interfere more than a pseudoword distractor as the activation of non-target lexical nodes should not influence lexical selection.

However, crucially, both competitive and non-competitive models fail to explain the reversal of the effects in a pseudoword context. That is, non-competitive models would still assume no influence of the word distractor on lexical selection of the target and as such would still predict no effect on naming latencies in a pseudoword context. Similarly, competitive models would still see a word as the strongest –and only- competitor, predicting longer naming latencies and more errors with words. Put differently, neither competitive nor non-competitive models have a built-in mechanism that deals with reversal of effects according to the context. Therefore, these models cannot accommodate the effects of Experiment 2 and 4 (pseudoword context). In contrast, the reversal of effects follows naturally from the monitoring account proposed in this paper as the monitor will adjust its criteria to the context of speech. Thus, including the monitor as a mechanism involved in fluent lexical access seems needed. This leaves us with two other options: competitive and non-competitive models subserved by a monitoring device. Further research is needed to determine whether the core assumption of lexical selection by competition is needed or not. However, based on the current results, we argue that this assumption is not necessary. That is, a non-competitive model with a monitor can explain these and other results (e.g., distractor frequency effect, semantic interference effect) equally well as a competitive model with a monitor, rendering a competitive mechanism unnecessary.

The data do not only have implications for models of lexical selection in general, they allow further specification of the response exclusion hypothesis as well. This hypothesis bears a strong resemblance to the monitoring hypothesis presented in this paper. Most importantly, according to the response exclusion hypothesis, responses to the distractor will enter a response buffer. Subsequently, they need to be checked for accuracy before the response to the picture enters the buffer. Thus, both the response exclusion hypothesis and the monitoring hypothesis assume a role for a control mechanism responsible for the effects in the PWI task. The results of the experiments reported in this paper invite the conclusion that the control mechanism could be the monitor. That is, it is active and adaptive as the effect of lexical status of the distractor changes over conditions. Second, the results of Experiments 2 and 4 indicate that the control mechanism needs time to perform its task. However, even though the mechanism is subject to time limitations, it does not fail totally when confronted with time pressure. Even under speed instructions, subjects were able to name the picture well on a majority of the trials. In sum, it seems as though the mechanism is a resource limited, active and adaptive mechanism. These are exactly the assumed features of the verbal self-monitor. Thus, one added value of this study is that it supports the idea that the exclusion mechanism is not an ad-hoc and task-specific mechanism but the monitor, a general mechanism implied in speech production.

However, note that the suggestion of an active device, sensitive to time pressure, is in contrast to previous formulations of the response exclusion hypothesis. Specifically, Mahon et al. (2007) explicitly argued against the involvement of the monitor, pleading for a more passive process instead of an active process of monitoring (p.524, footnote 4). They stated that representations at the response level have general properties of the concepts, like semantic category, and information concerning the source of the representation (i.e., picture or word), attached. Thus, task constraints determine parameters that are used to filter out representations that do not correspond to the target. This includes some sort of filtering, but is different from the active monitoring of the responses. The data reported in this paper however clearly support the involvement of the monitor, contrary to the mechanism Mahon et al. propose. Furthermore, Mahon et al.'s view is rather underspecified: it remains vague how the process of excluding the response actually proceeds. In our opinion, assigning a role for the monitor in the PWI task has far more explanatory power than the view proposed in Mahon et al: effects in this task and other production tasks can now be seen in a broader perspective.

These data also indicate a new role for the PWI task, namely as a tool to study the role of attentional processes that block out distracting information from speech production in general, and to study self-monitoring in particular. That is, in essence, the PWI task is a task that forces subjects to deal with distracting information, imposed by presenting a distractor. This way, the task offers a way to investigate the role of distracting information in natural speech. This is particularly interesting because despite the fact that in daily life, people are bombarded with various distracting stimuli, little is known about how this information is blocked from the language production system. One way this distracting information could be detected and deleted is through monitoring. Indeed, these data show that the monitor checks distracting information and removes it from the system, on the condition that it has enough time to do so. This conclusion is also supported by a recent paper by Dhooge and Hartsuiker (2011a). In that study, we had subjects name pictures with taboo and neutral distractors under speeded and standard instructions (i.e., emphasizing speed without sacrificing accuracy). Taboo intrusions occurred less frequently than neutral ones, indicative of a monitor filtering taboo errors out more often than neutral errors. Furthermore, naming latencies were longer when distractors were taboo words compared to neutral words. That finding suggests that the monitor adjusts the production process as a function of the micro context of a single trial. That is, the monitor would become more stringent in letting responses pass during a trial with a taboo context.

Similarly, the PWI task is usually not seen as a task to study monitoring. However, our data show that it could be an excellent tool to do so. After all, we found an equivalent of the lexical bias effect but the speeded version of the PWI task elicited many more errors than the SLIP task (e.g., Motley et al., 1982). Using the PWI task, the role of various monitoring criteria, imposed or not, can be investigated. For example, if it is claimed that the monitor uses a certain criterion in checking speech, it could be investigated whether the usage of this criterion is modulated by manipulations of time pressure and context. Furthermore, most studies concerned with the effect of monitoring on speech production focused on the pattern of errors caused by the interaction between speech production processes and the monitor. The issue is often whether particular speech error patterns plead for interactivity in production, or whether the data can be accounted for by a discrete production system, with a biased self-monitoring system (see Hartsuiker, 2006 for review). However, as it takes time to detect an error and remove it from the system, its effects should also be evident on naming latencies. Again, the PWI task provides an excellent tool to investigate this issue.

One final question relates to the paradoxical effect of phonologically related distractors. In a speeded PWI task, these distractors lead to more distractor naming errors than unrelated distractors (Starreveld & LaHeij, 1995), but at the same time lead to shorter naming latencies in a standard version of the PWI task (e.g., Schriefers et al.,1990). Judging from the intrusion error results, and following the logic followed in this paper, it could be concluded that these distractors are more difficult to detect as errors due to their similarity to the target. This would be the same logic as applied to semantic errors. However, it would also imply longer naming latencies with phonologically related distractors, mirroring the effect of semantic relatedness. Thus, the phonological facilitation effect cannot be explained by a monitoring account. One way to account for the effect was recently put forward by Mulatti and Coltheart (in press). These authors suggested that the effect of phonology has to be placed at an earlier level, before the response buffer in terms of the response exclusion. That is, if it is assumed that the target facilitates the formulation of a related distractor, this distractor would be ready for production earlier and can be excluded earlier as well. However, this account can still not explain the interaction between the semantic interference effect and the phonological facilitation effect (e.g., Damian & Martin, 1999; Starreveld & LaHeij, 1995). As the phonological facilitation effect arises earlier, these two effects should be independent from each other. Therefore, future formulations of the monitoring hypothesis should be able to account for this paradox.

In conclusion, four experiments have shown that the a control mechanism is implicated in the PWI task that is sensitive to lexical status of the distractor, that it is sensitive to time pressure and it adapts itself to the context of speech. All these features point to the verbal self-monitor. Therefore, these results indicate that the monitor should be taken into account when investigating the mechanisms of lexical selection.

References

Abdel Rahman, R., & Melinger, A. (2007). When bees hamper the production of honey: Lexical interference from associates in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 604-614.

Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. Language and Cognitive Processes, 24, 713-734.

Alario, F. X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *53A*, 741–764.

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.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). The CELEX lexical database. CD-ROM. Philadelphia, PA.

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, *49A*, 5-28.

Baddeley, A. D., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *Quarterly Journal of Experimental Psychology*, *51A*, 819-852.

Bates, D. M. (2007). *Linear mixed model implementation in lme4*. Unpublished manuscript, University of Wisconsin, Madison.

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

Bloem, I., & LaHeij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language, 48,* 468-488.

Bloem, I., van den Boogaard, S., & LaHeij, W. (2004). Semantic facilitation and semantic interference in language production: Further evidence for the conceptual selection model of lexical access. *Journal of Memory and Language, 51*, 307-323.

Brocklehurst, P. H., & Corley, M. (2011). Investigating the inner speech of people who stutter: Evidence for (and against) the covert repair theory. *Journal of Communication Disorders, 44*, 246-260.

Burt, J. S. (2002). Why do non-color words interfere with color naming? *Journal of Experimental Psychology: Human Perception and Performance*, *28,* 1019-1038.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177–208.

Caramazza, A., & Costa, A. (2000). The semantic interference effect in the picture–word interference paradigm: Does the response set matter*? Cognition*, *75*, B51-B64.

Costa, A., Alario, F. X., & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture–word interference paradigm. *Psychonomic Bulletin & Review, 12,* 125–131.

Catling, J. C., Dent, K., Johnston, R. A., & Balding, R. (2010). Age of acquisition, word frequency and picture-word interference. *Quarterly Journal of Experimental Psychology, 63*, 1304-1317.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25,* 345-361.

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

Dhooge, E., & Hartsuiker, R. J. (2010). The distractor frequency effect in picture-word interference: Evidence for response exclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 878-891.

Dhooge, E., & Hartsuiker, R. J. (2011a). How do speakers resist distraction? Evidence from a taboo picture-word interference task. *Psychological Science, 22,* 855-859.

Dhooge, E., & Hartsuiker, R. J. (2011b). The distractor frequency effect in a delayed picture-word interference task: further evidence for a late locus of distractor exclusion. *Psychonomic Bulletin & Review, 18,* 116-122.

Duyck, W., Anseel, F., Szmalec, A., Mestdagh P., Tavernier, A., & Hartsuiker, R. J. (2008). Improving accuracy in detecting acoustic onsets. *Journal of Experimental Psychology: Human Perception and Performance, 43*, 1317-1326.

Duyck, W., Desmet, T., Verbeke, L., & Bruysbaert, M. (2004). WordGen: A tool for word selection and non-word generation in Dutch, German, English, and French. *Behavior Research Methods, Instruments & Computers, 36,* 488-499.

Finkbeiner, M., & Caramazza, A. (2006). Now you see it, now you don't: On turning semantic interference into facilitation in a Stroop-like task. *Cortex, 42,* 790–796.

Ganuskchak, L. Y., & Schiller, N. O. (2006). Effects of time pressure on verbal self-monitoring: An ERP study. *Brain Research, 1125,* 104-115.

Ghering, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4,* 385-390.

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

Hartsuiker, R. J., Anton-Mendez, I., Roelstraete, B., & Costa, A. (2006). Spanish spoonerisms: A lexical bias effect in Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 949-953.

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

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology, 42*, 113-157.

Huettig, F., & Hartsuiker, R. J. (2010). Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Language and Cognitive Processes, 25,* 347-374.

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

Janssen, N., Schirm, W., Mahon, B. Z., & Caramazza, A. (2008). Semantic interference in a delayed naming task: Evidence for the response exclusion hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 249-256.

Jou, J., & Harris, R.J. (1992). The effect of divided attention on speech production. *Bulletin of the Psychonomic Society, 30,* 301-304.

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology, 77,* 576–588.

Lackner, J. R., & Tuller, B. H. (1979). Role of efference monitoring in the detection of self-produced speech errors. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing* (pp. 281-294). Hillsdale, N. J.: Erlbaum.

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

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

Lupker, S. J. (1979). The semantic nature of response competition in the picture–word interference task. *Memory & Cognition, 7,* 485–495.

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 503-533.

Miozzo, M., & Caramazza, A. (2003). When more is less: A counterintuitive effect of distractor frequency in picture–word interference paradigm. *Journal of Experimental Psychology: General, 132,* 228–252.

Mulatti., C., & Coltheart, M. (in press). Picture-word interference and the response-exclusion hypothesis. *Cortex.*

Nooteboom, S. (2005a). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech communications, 47,* 43-58.

Nooteboom, S. (2005b). Listening to oneself: Monitoring speech production. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, F. N. K. Wijnen (Eds.). *Phonological encoding and monitoring in normal and pathological speech.* Hove: Psychology Press.

Nooteboom, S., & 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.

Nozari, N., & Dell, G. S. (2009). More on lexical bias: How efficient can a lexical editor be? *Journal of Memory and Language, 60,* 291-307.

Nozari, N., Dell, G.S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology, 63,* 1-33.

Oomen, C. C. E., & Postma, A. (2002). Limitations in processing resources and speech monitoring. *Language and Cognitive Processes, 17,* 163-184.

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonetic similarity effect. *Cognition, 106,* 528-537.

Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & Cognition, 38,* 1147-1160.

Ozdemir, R., Roelofs, A., & Levelt, W. J. M. (2007). Perceptual uniqueness point effects in monitoring internal speech. *Cognition, 105,* 457-465.

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

Postma, A., & Kolk, H. H. J. (1992). The effects of noise masking and required accuracy on speech errors, disfluencies and self-repairs. *Journal of Speech and Hearing Research, 35,* 537-592.

Postma, A., & Kolk, H. H. J. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research, 36*, 472-487.

Postma, A., & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech, 39*, 375-392.

R Core Development Team (2009). *R: A language and environment for statistical computing.* Available from the R Foundation for Statistical computing Web site: http://www.R-project.org.

Riès, S., Janssen, N., Dufau, S., Alario, F.-X., Burle, B. (2011). General purpose monitoring during speech production. *Journal of Cognitive Neuroscience, 23,* 1419-1436.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142.

Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition*, *47,* 59–87.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, 64, 249-284.

Roelofs, A. (2001). Set size and repetition matter: Comment on Caramazza and Costa (2000). *Cognition, 80,* 283–290.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review, 110,* 88–125.

Roelofs, A., Piai, V., & Schriefers, H. (2011). Selective attention and distractor frequency in naming performance: Comment on Dhooge and Hartsuiker (2010). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1032-1038.

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture–word interference task. *Child Development*, *26*, 247–253.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production – picture word interference studies. *Journal of Memory and Language, 29,* 86-102.

Severens, E., Janssen, I., Kühn, S., Brass, M., & Hartsuiker, R. J. (in press). When the brain tames the tongue: Covert editing of inappropriate language. *Psychophysiology*.

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica, 119,* 159-187.

Starreveld, P. A., & LaHeij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 686–698.

Starreveld, P. A., & LaHeij, W. (1999). Word substitution errors in a speeded picture-word task. *American Journal of Psychology, 112,* 521-553.

Stevens, M., Lammertyn, J., Verbruggen, F., Vandierendonck, A. (2006). Tscope: A C library for programming cognitive experiments on the MS Windows platform. *Behavior Research Methods, 38*, 280-286.

Vigliocco, G., Lauer, M., Damian, M. F., & Levelt, W. J. M. (2002). Semantic and syntactic forces in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 46–58.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, *48*, 422-488

Vigliocco, G., Vinson, D. P., & Siri, S. (2005). Semantic similarity and grammatical class in naming actions. *Cognition*, *94*, B91-B100.

Ye, Z., & Zhou, X. (2009). Executive control in language processing. *Neuroscience and Biobehavioral Reviews*, *33*, 1168-1177.

Author Note

This research was supported by Research Foundation Flanders Grand FWO08/ASP/070 to the first author.

[i] Note that with random slopes, calculation of the p-value is not yet implemented in lme4.