Cognitive control during selection and repair in word production

Nazbanou Nozari[1,2], **Michael Freund**[1], **Bonnie Breining**[2], **Brenda Rapp**[2], and **Barry Gordon**[1,2]

[1]Department of Neurology, Johns Hopkins University School of Medicine

[2]Department of Cognitive Science, Johns Hopkins University

Abstract

Production of an intended word entails selection processes, in which first the lexical item and then its segments are selected among competitors, as well as processes that covertly or overtly repair dispreferred words. In two experiments, we studied the locus of the control processes involved in selection (*selection control*) and intercepting errors (*post-monitoring control*). Selection control was studied by manipulating the overlap (contextual similarity) in either semantics or in segments between two objects that participants repeatedly named. Post-monitoring control was examined by asking participants to reverse, within each block, the name of the two objects that were either semantically- or segmentally-related, thus suppressing a potent, but incorrect, response in favor of an alternative (reversal). Results showed robust costs of both contextual similarity (which increased with the degree of similarity between target and context) and reversal, but the two did not interact with one another. Analysis of individual differences revealed no reliable correlation between the cost of contextual similarity when pairs were semantically- or segmentally-related, suggesting stage-specific selection control processes. On the other hand, the cost of reversal was reliably correlated between semantically- and segmentally-related pairs, implying a different control process that is shared by both stages of production. Collectively, these results support a model in which selection control operates separately at lexical and segmental selection stages, but post-monitoring control operates on the segmentally-encoded outcome.

Keywords

Spoken word production; Written word production; Cognitive control; Executive function; Semantic blocking; Repair

Introduction

Producing a word requires mapping semantic features to lexical representations (lexical selection stage) and mapping those representations to segments consisting of phonemes in spoken production and graphemes in written production (segmental selection stage; e.g., Dell, Nozari, & Oppenheim, 2014). Little is known about the cognitive control processes that operate at each stage. Broadly defined, cognitive control refers to the ability to make

Corresponding Author: Nazbanou Nozari nozari@jhu.edu, Address: Neurology Department, Johns Hopkins University School of Medicine, 1629 Thames Street, Suite 350, Baltimore, MD 21231, Phone: 443-287-1712, Fax: 410-955-0188.

appropriate adjustments in perceptual selection, response biasing and online maintenance of information in order to accomplish the goal of a task (Botvinick, Braver, Barch, Carter, & Cohen, 2001). The need for cognitive control is assessed through signals from the monitoring system, which constantly evaluates the probability of achieving the task goal under the current circumstances. If this probability is deemed low, cognitive control is recruited (Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004). Deployment of control helps accomplish the task goal, although situations that demand high control often show a cost (e.g., slower performance), indicating that the recruitment of control is difficult and effortful (Ridderinkhof et al., 2004).

This study investigates two types of control processes that prevent errors in single-word production. The first is control required to resolve competition during lexical and/or segmental selection, which we call *selection control*. Selection control may operate at both stages of production, and its failure at each stage results in certain kinds of errors: if control fails during lexical selection, the most probable error is a semantic one (e.g., *dog* for *cat*). If control fails during segmental selection, a segmental error is produced (e.g., *mat* for *cat*). Because production of a word is more difficult in the context of similar items (e.g., Breining, Nozari, & Rapp, 2015; Damian, Vigliocco, & Levelt, 2001a; Kroll & Stewart, 1994; Schnur, 2014), the need for selection control can be assessed by manipulating the similarity between target and context and measuring the cost in performance (*contextual similarity cost*). Moreover, depending on the nature of similarity (semantic or segmental overlap), different stages of selection can be tapped.

The second type of control involves monitoring processes that intercept and correct a potent but inappropriate response that has either been selected or has a high chance of being selected. Speakers have the ability to correct errors covertly, before they emerge in overt speech (e.g., Nooteboom, 2010), or to internally replace less preferred responses with more appropriate ones (e.g., *woman* with *lady;* Levelt, 1983, 1989), a task that requires suppressing a strongly activated response and replacing it with a less potent one. We will refer to this type of control as *post-monitoring control*. We use this term not to refer to processes involved in detection of the error per se, but instead to refer to processes following the detection of an unsuitable response, specifically those involved in suppressing that response and replacing it with an alternative one. The need for this type of control can be assessed by asking speakers to reverse the names of familiar pictures (e.g., to produce *cat* whenever a picture of a *dog* is shown and vice versa) and measuring the cost in performance (*reversal cost*).

Throughout the paper, we use cost, defined as increase in response latencies (RTs) induced by either contextual similarity or reversal demands, as an indirect index of cognitive control, assuming that neurotypical speakers implement control to the best of their ability in response to demand for control. This study investigates two questions: (1) Does contextual similarity at both stages of selection increase the demand for selection control, and if so, is this control specific to each stage or shared between the two stages (Experiment 1)? (2) Does selection control interact with post-monitoring control (Experiment 2)? The latter result will be used to adjudicate between two models with different loci of operation for post-monitoring control.

Selection control

Most theories of word production assume that similar words compete for selection (but see Mahon, Costa, Peterson, Vargas, & Caramazza, 2007). Regardless of whether selection is directly affected by competitors (e.g., Howard, Nickels, Coltheart, & Cole-Virtue, 2006), or is indirectly influenced by how competitors shape the lexical network (e.g., Oppenheim, Dell, & Schwartz, 2010), there is unequivocal evidence that naming similar items in a set induces interference and requires selection control. Interference is well established in semantically-overlapping contexts (Belke, Meyer, & Damian, 2005; Crowther & Martin, 2014; Schnur, 2014; Damian et al., 2001; Kroll & Stewart, 1994; Schnur, 2014; Schnur et al., 2009). Similarly, in individuals with brain damage, the rate of semantic errors increases in a semantically-related context (e.g., Schnur, Schwartz, Brecher, & Hodgson, 2006). Oppenheim et al.'s (2010) model proposes a link between the interference induced by semantically-similar context and recruitment of cognitive control: when target selection becomes more difficult due to contextual similarity, a booster mechanism amplifies the activation of each word until one word becomes discernibly more active than others and can be selected. The time it takes for the booster to identify a clear winner reflects the cost of contextual similarity and the magnitude of the control required. This process is linked to the left ventrolateral prefrontal cortex, which has a role in biasing competition in both language production and comprehension (January, Trueswell, & Thompson-Schill, 2009; Kan & Thompson-Schill, 2004; Nozari & Thompson-Schill, 2013; Thompson-Schill et al., 1998).

Recently, we have argued that a segmentally-related context, defined as a context in which target words overlap in phonemes/letters, also interferes with production (Breining et al., 2015). Using a blocked cyclic naming paradigm, Breining and colleagues showed that segmental overlap led to robust interference when overlap was distributed unpredictably among words in a block (e.g., *pot, peg, leg, log, pig, pill*) compared to when the same items were named in contexts with low segmental overlap. While Breining and colleagues' results indicate increased demand for selection control when words in a naming set had unpredictable segmental overlap, it is not clear that overlap in segments, especially if the overlap is predictable, necessarily leads to interference (see Goldrick, Folk, & Rapp, 2010 for a review of facilitatory effects of phonological overlap; but see O'Seaghdha & Marin, 2000; Rogers & Storkel, 1998; Sevald & Dell, 1994; Sullivan & Riffel, 1999; Wheeldon, 2003 for evidence of interference). Note that when all or the majority of words in a block share onset segments, it is typical to observe facilitation, an effect that has been attributed to strategic preparation (Damian & Bowers, 2003; Meyer, 1990; O'Séaghdha & Frazer, 2014; Roelofs, 1999; Shen, Damian, & Stadthagen-Gonzalez, 2013; but see Belke & Meyer, 2007 for hints of interference). Overlap in the initial segments is a special case, because it means that before each picture appears, the speakers can prepare the immediate segment with 100% certainty. In the non-initial segment overlap, on the other hand, the immediate segment to be produced is uncertain. Since segmental encoding takes place from left to right (Sevald & Dell, 1994), in the absence of certainty about the identity of the initial segment, knowledge about later segments may not be particularly useful. Thus, Experiment 1 was designed to assess if demand for selection control is indeed increased as a function of similarity in both stages of production, even when segmental overlap (in non-onset positions) is completely

predictable. In a blocked cyclic naming paradigm, participants repeatedly produced one of two words that were related either semantically (e.g., *hat, wig*) or segmentally (e.g., *hut, nut*), and the cost of relatedness was assessed against production of the same words in an unrelated context. For segmental overlap we considered both pairs that had overlap in rhyme and those that overlapped only in their last phoneme/letter. If similarity creates interference at both stages of production, one would expect interference in both semantically- and segmentally-overlapping contexts, as long as strategic preparation of segments is minimized by removing onset overlap.

If there is a robust cost for both semantic and segmental overlap, we can test if the two are correlated or not. A positive and reliable correlation implies a shared component, while lack of such correlation is consistent with stage-specific processes. The evidence for stage-specificity of semantic-lexical and lexical-segmental mapping comes primarily from the study of speech errors in neurotypical speakers and individuals with aphasia. Errors that arise from these two stages of mapping show different characteristics (see Dell et al., 2014 for a review), and that damage in each stage creates a different error profile (e.g., Dell, Schwartz, Martin, Saffran, & Gagnon, 1997). This stage-specific processing is also reflected in the process of error detection: Nozari and colleagues showed that the ability to detect semantic or phonological errors could be selectively impaired in aphasia, and such impairment reflected the extent of damage to the specific stage at which the error was generated (Nozari et al., 2011). If stage-specific processing relies on stage-specific control, then no correlation between costs – as indices of selection control – should be observed between semantically- and segmentally-related pairs. However, if both stages are controlled by a common selection control process, such a correlation would be expected.

Given the variability in experimental results regarding the consequences of segmental overlap, if interference is found in Experiment 1, it would be important to evaluate its reliability. To this end, we made three provisions. First, Experiment 1 was conducted in both spoken and written modalities. While spoken and written production certainly interact (see Rastle & Brysbaert, 2006, for a review), they can operate as largely independent systems (e.g., Bonin, Peereman, & Fayol, 2001; Miceli, Benvegnu, Capasso, & Caramazza, 1997; Rapp, Benzing, & Caramazza, 1997; Rapp & Fischer-Baum, 2014; Rapp, Fischer-Baum, & Miozzo, 2015). Specifically, conditions that lead to facilitation due to overlap in graphemes may operate independently of phonological overlap, at least in English (e.g., Shen et al., 2013). As such, replication of the effect in two modalities would increase our confidence in claiming that selection control is indeed necessary in segmentally-overlapping contexts. Second, each related condition contained items with high and low degrees of similarity. For segmental overlap, high similarity was defined as overlap in two segments (e.g., *hut, nut*), and low as overlap in only one (e.g., *cup, map*). For semantic similarity, subjective norms were collected and used to divide pairs into high and low similarity groups (e.g., *hat, wig* (high) vs. *heel, suit* (low)). If contextual similarity drives the interference, greater similarity should cause larger interference effects. Previous research examining contextual similarity costs have generally used a binary manipulation (related vs. mixed context), without analyzing the degree of similarity within the related condition (e.g., Breining et al., 2015). It thus remains possible that factors other than target-context similarity, e.g., specific strategies adopted by speakers, may have driven prior interference effects. Manipulation of degree of

similarity and tracking its effect on observed interference serves as a further test of the claim that observing greater interference in related compared to mixed contexts is a direct consequence of target-context similarity. Third, Experiment 2, although primarily designed for a different purpose (see below), additionally tested the replicability of Experiment 1's spoken production findings in a separate group of participants.

Although using cycles of only two words is not routine in cyclic naming paradigms, it might be preferable to using five or six words per block, as using only two items removes any potential differences across experimental conditions in working memory demands that could influence repeated naming of a limited set of pictures in cycles (Crowther & Martin, 2014). To evaluate that the paradigm does not behave substantially differently from those that include a larger item set per cycle, we also included an initial-overlap condition, which has consistently yielded strategic facilitation in prior studies (e.g., O'Séaghdha & Frazer, 2014; Roelofs, 1999; Shen et al., 2013). Results of this condition will be reported as a quick check, but will not be the focus of the analyses, since our main claims will not concern strategic facilitatory processes. In sum, results of Experiment 1 will show if there is a robust need for selection control in each stage of word selection (lexical and segmental), by looking at the contextual similarity cost. Moreover, a correlation between the costs for lexical and segmental selection would point to a shared mechanism, while the absence of such correlation would be consistent with separate selection-control mechanisms.

Post-monitoring control

Although selection is critical to production, not all selected words are suitable for production. Some are errors, and some are inappropriate for production in certain social situations or for addressing a certain audience. Speakers can monitor their speech and change the words that have been selected or are the strongest candidates for selection because of their high activation. Regardless of how detection of an error or an inappropriate word is accomplished (See Hickok, 2012; Nozari, Dell, & Schwartz, 2011; Pickering & Garrod, 2013 for different proposals), once a word is deemed unsuitable for production, it can be overridden by a newly-selected response. This may be done covertly, although some errors do surface in overt speech (e.g., Nooteboom, 2010; Postma, 2000), reflecting an inherent difficulty in suppressing a strong selection candidate and replacing it with another response that is less potent. Overcoming this difficulty requires cognitive control, especially for covert repairs, which occur before the dispreferred response is overtly produced. Several psychological paradigms are relevant here. Below, we review these paradigms, discuss their strengths and weaknesses and propose a paradigm that, in our opinion, most closely captures the cognitive operations behind covert repairs.

In terms of cognitive control, overriding a selected response with an alternative one is very similar to the core ability captured by the Stroop task (Stroop, 1935), in which the spoken representation of a written color word is strongly activated but must be suppressed in order to select the ink color's name as the correct response. The magnitude of the Stroop effect is measured as the difference in the accuracy or RTs between congruent (the word *red* printed in red ink) and incongruent trials (the word *red* printed in blue ink). Classic Stroop is sensitive to both semantic and segmental overlap between the printed word and the ink color

to be named. For example, Klein (1964) showed that the magnitude of the Stroop effect was larger when the ink color *red* was to be named for the word *yellow* or the word *lemon* which indirectly activates *yellow*, compared to a neutral word like *put* which is not semantically associated with color words. On the other hand, Coltheart, Woollams, Kinoshita, and Perry (1999) showed that phonological similarity decreased the magnitude of Stroop: naming the color red was faster when the printed word was *rat* compared to the unrelated word *kit* (see also Navarrete & Costa, 2005). Overlap in the coda (e.g., *pod*) also led to facilitation, although smaller in magnitude than that of the onset overlap. Similar semantic interference and phonological facilitation have been reported in picture-word interference studies where participants must ignore a written word and name a picture (e.g., Schriefers, Meyer, & Levelt, 1990).

While this line of research clearly shows the influence of both semantic and segmental overlap on the ability to suppress a potent response in favor of a new one, the classic Stroop task differs from speech monitoring in an important way: it involves competition between goals (reading the orthographic form vs. naming the ink color). Biasing the processing towards the correct goal is a critical aspect of Stroop, and is often captured by a "task node", a surrogate for the implementation of top-down control, in computational models of the task (e.g., Botvinick et al., 2001; Roelofs, 2003). During speaking, on the other hand, competition is not between task goals, but between representations within the same task (i.e., naming). Tydgat, Diependaele, Hartsuiker, and Pickering (2012) approached the issue of similarity between a to-be-suppressed response and its replacement using a different paradigm, which was closer to speech monitoring and repair in that it involved no reading. Participants attempted to name a picture, which was quickly replaced by another picture that was to be named instead. The second picture could be semantically related to the first picture, share the onset with its name, or be unrelated. They found facilitation for naming semantically-related pictures (but see Hartsuiker, Pickering, & De Jong, 2005), as well as interference for naming phonologically-related pictures when the name of the first picture was at least partially articulated.

The findings of facilitation with semantic overlap and interference with segmental overlap of Tygdat and colleagues are in direct contrast with those obtained from Stroop tasks. From the standpoint of approximating monitoring and repair processes in spoken production, Tygdat et al.'s (2012) study has the advantage of avoiding the activation of the reading route. On the other hand, Stroop studies have the advantage of simultaneously activating both responses. In monitoring everyday speech (Nozari et al., 2011), this simultaneous as opposed to serial activation is critical to the generation of the error signal that is presumed to trigger the processes responsible for intercepting the dispreferred response (stop signal) and replacing it with a new one (repair). This stop signal is externally generated in Tygdat and colleagues' paradigm (see Verbruggen & Logan, 2008 for differences between externally- and internally-generated stop signals). Another feature of Tygdat and colleagues' design is that participants do not know the identity of the target (i.e., the name of the second picture) when they start naming the first picture, which is rarely the case with in everyday speech when individuals repair an error with a correct response. This feature has critical consequences, especially for the facilitation observed in the semantically-related condition, and we will return to it in the General Discussion.

A third group of studies has used paradigms in which two pictures are simultaneously presented, but only one is to be named while the other is to be ignored (Picture-Picture Interference paradigm; PPI). The general finding in these paradigms is no effect of semantically-related distractors[1] (e.g., Damian & Bowers, 2003; Navarrete & Costa, 2005) and a facilitatory effect of phonologically-related distractors (e.g., Meyer & Damian, 2007; Navarrete & Costa, 2005; Roelofs, 2008), although the generality of the latter finding is debated (Bloem, van den Boogaard, & La Heij, 2004; Jescheniak et al., 2009; Oppermann, Jescheniak, & Görges, 2014). This design has the advantages of avoiding reading while potentially activating both responses. However, as the distractor is not intended to be produced, there is a good chance that it is not the more potent response, a situation which again differs from intercepting an error in speech monitoring.

The goal of Experiment 2 in the current study was to combine the positive features of the above designs, while avoiding as much as possible those features that made the tasks different from speech monitoring. Similar to Stroop tasks, the to-be-ignored response was the more potent one; similar to Tydgat et al. (2012), reading was removed; and similar to PPI tasks, both responses were simultaneously active and competing for selection with no external stop-signal. As in Experiment 1, each block contained two pictures, only one of which was named on each trial. However, in Experiment 2, half the blocks were "reversed" blocks, in which participants were asked to monitor their responses and call each picture by the name of the other picture in the block. Thus, in a reversed block containing the pictures of *nut* and *hut*, upon seeing the picture of the hut, they would produce the word *nut* and vice versa. Evidence from picture-picture interference paradigms (e.g., Meyer & Damian, 2007) has shown that the phonological form of a to-be-ignored picture-name can affect naming of a target picture, showing automatic activation of picture names down to the level of phonology. Given this finding, we assume that upon seeing the picture of a hut, the word *hut* is activated and competing with the target *nut* for selection. However, because of task demands, post-monitoring control processes must suppress *hut* and instead produce *nut*. The magnitude of post-monitoring control in this task is reflected in the cost of reversal, defined as RT(reversed) – RT(straight). We will assess if this cost interacts with the contextual similarity cost, defined as RT(related) – RT(mixed), which indexes selection control.

We propose two possible models for the interaction between selection and post-monitoring control (Figure 1). The left panel (Model 1) shows a model in which post-monitoring control is enforced only on the production output. The right panel (Model 2) shows a competing model in which post-monitoring control is enforced at each selection level. While both models predict independent contextual similarity costs for semantically- and segmentally-related pairs because selection control is enforced independently at the level of lexical and segmental selection, the two models make different predictions regarding the sensitivity of the reversal cost to contextual similarity; we examine two sets of predictions: (1) Model 1 predicts that reversal costs should be comparable in size for semantically- or segmentally-related pairs, because monitoring control happens at a later stage than the one in which contextual similarity influences performance. At a later stage, reversal only adds a constant

---

[1]Note that this is a finding from picture-picture interference paradigms, and not from the picture-word interference paradigms in which a semantically-related word distractor often creates interference.

to the similarity cost. Statistically, this possibility would manifest as the absence of a reliable interaction between contextual similarity and reversal costs. Model 2, on the other hand, predicts an interaction between reversal and contextual costs, as they happen interdependently. We test these predictions in Experiment 2. (2) Model 1 predicts that, at the level of individual participants, the reversal cost should be correlated between semantically- and segmentally-related pairs, because post-monitoring control for both types of pairs is applied at a single stage after similarity has affected selection. Model 2, on the other hand, predicts no correlation between reversal costs for semantically- and segmentally-related pairs.

To summarize, we test the robustness of the need for selection control in lexical and segmental selection and the stage-specificity of such control in Experiment 1, attempt to replicate the same findings in Experiment 2, and test the relationship between selection and post-monitoring control in Experiment 2.

Experiment 1

Experiment 1 tested whether the demand for selection control is robust at both lexical and segmental stages of selection and whether it increases as a function of similarity in both stages of word production.

Methods

Participants—Thirty-two native English speakers (18 women; mean age = 21.5 years) participated for payment or course credit. Data from one participant was lost due to technical problems.

Materials—Three related conditions (semantic, initial-overlap, and final-overlap) were created, each containing eight pairs of monosyllabic words that were related in their semantics (e.g., *hat/wig*), in their final segments (e.g., *cup/map*), or in their initial segments (e.g., *pen/pot*). Half of the pairs in the final-overlap condition shared only one segment, i.e., the coda (low-overlap; e.g., *cup/map*), and the other half shared two segments, i.e., the rhyme (high-overlap; e.g., *hut/nut*). The same was true for the initial overlap (e.g., *pen/pot* and *chip/chin*). Semantic pairs were divided into high and low similarity according to the ratings of 25 independent raters at Amazon's Mechanical Turk, who viewed pairs of pictures and rated their similarity on a scale of 1 to 7 (M_{High} = 5.36, SE = 0.15; M_{Low} = 3.84, SE = 0.17; non-parametric z = 5.99, p <0.001). Mixed pairs were constructed by pseudo-randomly reshuffling the words in the pairs, such that the two words in the new (mixed) pairs had negligible semantic similarity (M_{mixed}= 1.65, SE = 0.09; non-parametric z between low similarity and mixed pairs = 8.89, p <0.001) and no segmental overlap, but were matched in frequency and number of segments to the words in the related pairs. This created a total of 48 (24 related and 24 control) pairs. Forty-eight 300x300-pixel black and white line-drawings corresponding to the 48 words in the experiment were selected from the IPNP corpus (Szekely et al., 2004) and Google images. Four lists were created with pseudo-randomized order of picture pairs, such that the same word or related condition was never immediately repeated.

Procedures—The experiment was run in E-Prime 2.0 software (Psychology Software Tools, Pittsburg, PA). Pictures were displayed at the center of a 15 x 12 inch Dell monitor approximately 25 inches in front of the participants. Response times (RTs) for spoken responses were registered using an Audio-Technica microphone connected to the E-Prime's SR-BOX. RTs for written responses were collected using a Wacom Bamboo graphic tablet on which participants wrote their responses. Both spoken and written responses were recorded for later transcription and error identification.

Participants were randomly assigned to one of the four lists and completed two sessions, one spoken and one written, at least three days apart. The same list was used for both modalities in each participant. Each block consisted of one pair of items. For each pair in a block, they first saw the two pictures along with written labels (e.g., *hut/nut*), named each, received feedback, and then completed four practice trials. As instructed, on the next 16 trials, they named one of the two pictures at a time as quickly and accurately as they could (8 presentations of each picture in pseudo-randomized order so that the same picture never appeared more than twice consecutively). In the spoken version, each trial began with fixation cross presented at the center of the screen for 700 ms. The stimulus was then displayed for 2000 ms, or until a response was made. In the written version, the picture was replaced by an image of participant's handwriting as soon as they started to write. They had 2000 ms (determined by piloting) to finish their written response.

Results and Discussion

Overall error rates were 4% and <1% in the spoken and written modalities respectively (Table 1). Wilcoxon signed-rank test revealed no reliable differences between error rates in related and their corresponding mixed conditions (all tests had P>0.1). Error trials were removed, along with trials with RTs more extreme than 3SD from the mean of each participant's RT distribution. In addition, RTs shorter than 200 ms were discarded as manual checking of the acoustic waves showed that these were due to premature triggering of the microphone and did not reflect true decision processes in picture naming. This resulted in the exclusion of 8% of the data in the spoken version, and 4% of the data in the written version. The remaining RTs were log-transformed and inspection of the transformed data using QQ plots revealed an acceptable approximation to normal distribution. All analyses were conducted on the log-transformed data using multilevel mixed models with random effects in R version 3.1.0, with the lmerTest package.

Three models were constructed: Semantic, Final segment and Initial segment. Each model included Context (related vs. mixed), Modality (spoken vs. written), and Degree of similarity (high vs. low), two- and three-way interactions between those, and the control variable order (whether spoken or written modality was completed first) as fixed effects. The Context variable tests the main question of interest: is naming pictures more difficult in a related compared to a mixed context? The answer is directly comparable to the past reports of the inhibitory effect of semantically-related context (e.g., Schnur et al., 2006, 2009) and segmentally-related context (Breining et al., 2015). The Degree of similarity variable is included as a second check, to show that the interference between related and unrelated contexts relates directly to the degree of similarity between context and target and not any

other potential differences between the related and mixed conditions. All items are coded as either High or Low, regardless of whether they appear in the related or mixed conditions. This coding allows for testing whether there are main differences between items in the two groups. No reliable main effect would then allow us to test if the appearance of items, which did not differ in their basic properties, in related and mixed conditions will vary differentially, i.e., whether the High similarity items would show an even greater effect of Context than the Low similarity ones. This is tested in the interaction term between Context and Degree of similarity. The Modality variable and its interaction terms test the main effect of modality of RTs, as well as the potential differential sensitivity of the effects of Context and Degree of similarity to spoken vs. written production. Finally, the models include a covariate, Order, which codes whether participants were first exposed to the spoken or the written modality. For both Experiments we attempted to implement a full random structure in the model following recommendations of Barr, Levy, Scheepers, and Tily (2013). However, due to the lack of convergence in some models, the random slopes over items were dropped. The final model had random intercepts for both subjects and items, as well as the full random slope structure of the fixed effect structure of interest (context, modality and degree of similarity) over subjects. This architecture was kept constant across all models. All variables were centered, and categorical variables were contrast coded as –0.5 and 0.5.

Before we discuss the results of semantic and final-segment overlap, we report the results of initial segmental overlap, as a check of the paradigm. Recall from the Introduction that past studies have reported a facilitatory effect of context when the overlap is in the initial segments. In agreement with past studies, onset overlap led to significant facilitation in both spoken (M = 359; between-subject SE = 9.47 ms in the related and M = 368; between-subject SE = 9.29 ms in the mixed condition) and written modalities (M = 452; between-subject SE =17.62 ms in the related and M = 508; between-subject SE = 18.75 ms in the mixed condition). Shared onsets reliably facilitated production (t= −9.36, p <0.001), and caused significantly more facilitation in written than spoken production, reflected in the interaction term between context and modality (t = 12.41, p<0.001). This shows that in this respect the paradigm produced onset facilitation effects comparable to previously reported experiments with cycles containing more items. We now focus on the critical conditions in the experiment which potentially create interference and demand control for selection, namely semantic and final segment overlap.

Figure 2 shows the average RTs (+SE) for different conditions, collapsed over different degrees of similarity, for spoken and written sessions (See Table 2 for RTs±SE in each condition). Both semantic and final segment overlap show a contextual similarity cost. These effects were statistically tested in separate models for semantic and final segment overlap.

Semantic model—Table 3 shows the full results of this analysis. Not surprisingly, responses were faster in the spoken than written modality (modality; t = 10.7, p<0.001). Pictures were also named more quickly if participants had first completed the written modality (order; t = −3.00, p = 0.006). There were no reliable differences between RT's for naming pictures in high- and low-similarity conditions when collapsed over related and mixed context, reflecting the similar frequency and length of these items (Degree of similarity; t=0.91, p = 0.37). Critically, the model showed a reliable effect of context, with

related context significantly slowing production (Context; t = 2.28, p = 0.032). High-similarity had a significantly stronger effect in slowing production than low-similarity, as evidenced by an interaction between context and degree of similarity (t=-4.08; p<0.001). Modality did not interact with either the effect of context or the influence of degree of similarity over context, tested by the two- and three-way interactions between context and modality, and context, degree of similarity, and modality, respectively.

To further explore the effects of semantic overlap in each modality, two post-hoc models were built, one for each modality. The structure of the models was the same as described above, except that the Modality variable was removed. The Spoken model showed a robust effect of Context (t = 2.47, p=0.038), as well as a significant interaction between Context and Degree of similarity (t = −3.77, p <0.001), corrected for multiple comparisons. The written model also revealed a reliable effect of Context (t = 2.44, p = 0.018) and a significant interaction between Context and Degree of similarity (t = −1.99, p = 0.047). These post-hoc tests demonstrated that, regardless of modality, related context interfered with picture naming and such interference were stronger when the target and context were more similar.

Final segment overlap model—Table 4 shows the full results of this analysis. As in the semantic model, production was faster in the spoken modality (t=12.76, p<0.001), and as before, there were no reliable differences between the RTs for pictures in high- and low-similarity conditions (t=1.53, p =0.14). Critically, the model showed a reliable effect of Context, with related context significantly slowing production (t=4.95, p<0.001). Again, the high-similarity context had a significantly stronger effect than low similarity is slowing production, as evidenced by an interaction between Context and Degree of similarity (t=-2.30, p=0.021). There was also a marginal three-way interaction between Context, Degree of similarity, and Modality (t=1.78, p = 0.075). The latter tests whether contextual similarity is equally sensitive to the degree of similarity in spoken and written production, and the marginal interaction suggests that it may not be. As can be seen in Table 2, the cost of segmental overlap, operationalized as RT(related) – RT(mixed) is 15 ms for High and 4 ms for Low similarity pairs in spoken production, revealing an 11ms difference. The same comparison shows a 12 ms vs. 9 ms cost for High and Low overlap in written production, a small difference of 3 ms. It is thus possible that written production is not sensitive to the degree of similarity between target and context as spoken production is. Post-hoc tests examine this possibility.

Similar to the semantic model, we built two post-hoc models to test the effect of contextual similarity separately in spoken and written production. The Spoken model showed a robust effect of Context (t = 4.33, p<0.001), as well as a significant interaction between Context and Degree of similarity (t=-2.76, p = 0.006). The written model also revealed a reliable effect of Context (t = 2.47, p = 0.034) but no reliable interaction between Context and Degree of similarity (t = −0.39, p = 0.70). The results of the post-hoc tests converged with the main model: in both spoken and written modalities, a segmentally-related context caused interference in picture naming. However, only spoken production showed reliable sensitivity to the degree of similarity between the target and context.

In summary, interference was found as a function of both semantic and segmental contextual relatedness in both spoken and written modalities. There was a general tendency for the magnitude of interference to increase as the degree of similarity between the words in a pair increased. One exception was segmental overlap in written production, which was less sensitive to degree of similarity than spoken production. This difference manifested as a non-significant three-way interaction between context, degree of similarity and modality in the segmental model, and was supported by the post-hoc tests. The greater interference for high (rhyme) compared to low (coda) overlap in the spoken modality may reflect spoken production's greater sensitivity to syllabic structure, but may also result from written production being slower and more serial. Critical to the goals of this study, however, both modalities showed interference as a consequence of both semantic and segmental overlap. These results provide reliable evidence that the need for selection control increases as a function of contextual similarity in both stages of selection.

Experiment 2

Experiment 1 showed a robust cost for semantic and segmental contextual similarity. Experiment 2 evaluated whether this cost interacted with the reversal cost indexing post-monitoring control.

Methods

Participants—Thirty-two native English speakers (23 women; mean age = 20.16 years) participated for payment or course credit. None had participated in Experiment 1.

Materials—The same materials as Experiment 1 were used. However, the number of experimental trials for each picture pair in the straight naming condition was reduced from 16 to 8 (four presentations of each picture in a pair). Instead, a reversed phase was added in which participants switched the name of the two pictures in a pair. Thus, each picture was named eight times in the straight and eight times in the reversed phase, but the total number of trials remained the same as in Experiment 1. Order of appearance of words in straight or reversed phases was counterbalanced across participants. Only the spoken modality was tested.

Procedures—Procedures were similar to those used for the spoken version in Experiment 1. At the beginning of the block of trials for each picture pair , the word "STRAIGHT" or "REVERSED" informed the participant of whether pictures names were to be reversed or not, and participants orally confirmed this to the experimenter before starting each block.

Results and Discussion—The overall error rate was 5%, with no reliable differences between error rates in related contexts and their corresponding mixed contexts, tested by Wilcoxon signed-rank test (all tests had P>0.1). With the exclusion of trials with RT<200 ms and those more extreme than 3SD from each participant's mean, 9% of responses were excluded from the analyses. The rest of the RTs were log-transformed for analysis using a multi-level model with mixed effects. The upper panel in Figure 3 shows the average RTs (+SE) for all contexts, separately for straight and reversed phases (upper panel). As can be seen, the contextual cost is present in the reversed as well as the straight phase. The lower

panel plots the reversal (+SE) for each context, with no evidence of reduction in the reversal cost in the related contexts.

The reliability of the reversal cost (i.e., need for post-monitoring control) as well as its interaction with contextual similarity were statistically tested in a model with the following fixed effects: Context (related vs. mixed), Relation (semantic vs. segmental), Phase (straight vs. reversed), and two- and three-way interactions between them[2]. Instead of separate models for semantic and segmental overlap, we ran a single model on both datasets. The reason for doing so is as follows: in Experiment 1 our main interest was testing the contextual similarity effect, and especially to test if our earlier claim that segmental overlap generally causes interference during picture naming (Breining et al., 2015) was supported. To this end, we conducted detailed analyses in two modalities, including post-hoc tests, to show that both semantic and non-initial segmental overlap between target and context led to interference, and in the majority of cases, this interference increased as the similarity between the target and context increased. Experiment 2 pursues a different goal. Now that we have confirmed that context at each stage of selection creates interference and demand for selection control, we ask if post-monitoring control interacts with selection control. Any kind of interaction between post-monitoring control (indexed by the variable Phase in the model) and the variables indexing selection control would show the sensitivity of post-monitoring control to selection mechanisms that happen at each stage of selection, and point to a model in which post-monitoring control is implemented at the same stage as selection. This can be tested by examining the two-way interaction between Phase and Context, or the three-way interaction between Phase, Context, and Relation. The former would show that post-monitoring control is affected by the dynamics of selection control at either stage of selection. The latter would demonstrate that post-monitoring control is differentially affected by selection control at the stage of lexical versus segmental selection. An absence of a reliable interaction on either term would mean that post-monitoring control is unaffected by selection control, regardless of the selection stage. A single model that accommodates these terms has the highest statistical power to test the sensitivity of post-monitoring control to selection control. This model was constructed following the same rules for the inclusion of random effects as in Experiment 1.

Similar to Experiment 1, we first examined the effect of initial-segment overlap in a separate model. Onset facilitation disappeared in Experiment 2. For onset overlapping pairs, there were no reliable differences between mean RTs in the straight (390±6 in related vs. 390±7 in mixed) or reversed (459±9 in related vs. 461±7) phases (main effect of Context: t=-0.12, p = 0.91; main effect of phase: t = − 9.97, p < 0.001; no reliable interaction between context and Phase). Disappearance of onset facilitation is in line with claims of the effect being strategic and not reflecting stable dynamics of the production system (e.g., O'Séaghdha & Frazer, 2014; Roelofs, 1999). As before, we focus the analyses on conditions that led to interference in Experiment 1: semantic and final segment overlap.

---

[2]We did not include degree of similarity, because the critical predictions of Experiment 2 do not depend on that variable. To ensure that this exclusion did not significantly change the model fit, we tested the reported model against a full model including degree-of-similarity and all its interactions. The change in fit was not significant ($\chi 2 = 11.7$, p = 0.16).

Table 5 presents the results of this analysis. Production was reliably slower in the related context (Context; t = 2.1, p = 0.04), with no significant interaction between Context and Relation (t=-0.8, p =0.40). There was also a robust effect of Phase, with reversed responses significantly slower than straight responses (t=-13.9, p<0.001). There was, however, no evidence of interaction between the effect of context and the effect of reversal, either in the interaction between Context and Phase (t=0.1, p = 0.96), or in the three-way interaction between context, relation, and phase (t=-1.3, p = 0.20).

In summary, Experiment 2 replicated the reliable contextual similarity found in Experiment 1. An unexpected finding of Experiment 2 was the disappearance of the onset facilitation effect which was found in Experiment 1 in keeping with findings of past studies which manipulated onset similarity in cyclic naming paradigms. Most likely, the fact that on each block participants had to remember whether the block was straight or reversed occupied the conscious attentional resources that would have otherwise been dedicated to preparing the initial shared segment. This explanation is in line with the onset facilitation effect being considered strategic and attentional, and not part of the internal dynamics of the language production system (O'Séaghdha & Frazer, 2014). Critically, for the hypotheses of the study, this experiment also showed a reliable need for post-monitoring control (reversal cost), which did not interact with selection control. The results showed that contextual similarity was present in both straight and reversed phases with comparable magnitude, as apparent in the lack of a significant interaction between context and phase. A corollary to this was that reversal costs were also similar in magnitude between related and mixed contexts for both semantically- and segmentally-related pairs. These results support the first prediction of Model 1, in which post-monitoring conflict operates after both selection stages have completed. The second prediction of this model is tested below using an analysis of individual differences.

Analysis of individual differences

To reiterate, the contextual similarity cost was defined as RT(related) – RT(mixed), which indexes selection control. This cost was calculated separately for semantically- and segmentally-related pairs. The correlation between contextual similarity cost for these semantically- and segmentally-related pairs was small and unreliable (Pearson's r = 0.03, p = 0.89) in Experiment 1. The same was true for Experiment 2 (Pearson's r = –0.06, p = 0.73; Figure 4, upper panel). Combining the two datasets, the overall correlation between semantically- and segmentally-related costs remains close to zero (Person's r = –0.03, p = 0.83)[3]. This is unlikely to be caused by lack of statistical power. While there is no a priori effect size, based on which a sample size can be calculated, for detecting a medium-sized effect (q = 0.6), with α = 0.05, and power = .80, detection of a within-subject correlation requires N = 19 (38 data points). With 41 participants, it is possible to refute the null hypothesis for the aforementioned effect size with a power = 0.99. Thus, statistical power does not seem to be an issue. A second limiting factor in the interpretation of a null result is the internal consistency of the measures, which is infamously low for many psychological

---

[3]To assess the correlation without the influence of potential outliers, we also calculated the non-parametric Spearman's rank-ordered correlation index, which was –0.08, p = 0.56.

constructs, such as difference scores (Redick et al., 2013). To measure the internal consistency of the semantically- and segmentally-related costs, we first split the trials for each word in related and mixed conditions into odd and even, calculated an average for each, and then calculated a cost by subtracting the average production time in the related minus mixed condition, for each word in each participant. These costs were then averaged for all words belonging to the same type of context (semantic or segmental), creating four measures in each participant: semantic cost in odd trials, semantic cost in even trials, segmental cost in odd trials, and segmental cost in even trials. From these, we calculated the internal consistency of semantic and segmental costs, by correlating each type of cost on even trials with its corresponding cost on odd trials. The correlation index (r) derived from this

procedure was then used to calculated the Spearman-Brown split-half reliability$\left(\dfrac{2r}{1+r}\right)$. This procedure returned a split-half reliability of 0.63 for segmental costs and 0.44 for semantic costs. The intermediate values of these indices invite caution in interpreting the null result across the two costs. However, at this point the weight of the evidence is more compatible with independent selection control processes for lexical and segmental selection. Thus, the Models that are proposed to evaluate the relationship between selection and monitoring control have stage-specific selection control mechanisms. The critical difference between these models is the locus of post-monitoring control. The magnitude of post-monitoring control was indexed by the cost of reversal, defined as RT(reversed) – RT(straight). The locus of post-monitoring control in Model 1 is at the output level, thus this model predicts a positive and reliable correlation between the reversal cost for semantically- and segmentally-related words, because the same control mechanism is applied to the output of selection processes, where differences in context are no longer pertinent. On the other hand, Model 2 predicts no such correlation, as the post-monitoring control is enforced separately at each selection level, similar to the selection control. The lower panel of Figure 4 shows the correlation between reversal cost for pairs that overlapped in semantics and final segments. This correlation was positive and reliable (Pearson's r = 0.57, p = 0.001)[4], supporting the prediction of Model 1. Together with the results of the prior analyses, the data support both predictions of a model in which post-monitoring control operates on the same output after both stages of selection have been completed.

General Discussion

In two experiments, we studied two types of control processes involved in word production. Selection control was defined as the speaker's ability to suppress the interference from semantically or segmentally similar contexts and was measured by the contextual similarity cost. Post-monitoring control was defined as the speaker's ability to suppress a potent but dispreferred response before it surfaced in overt production and was measured by the reversal cost. Experiment 1 tested selection control, showing that overlap in either semantic or final segments between target and context led to reliable interference. The presence of interference in the segmentally-overlapping context extends the findings of Breining et al. (2015) to situations where (non-onset) segmental overlap is completely predictable. To

---

[4]The non-paramteric Spearman's rank-ordered correlation index for this correlation was 0.52, p = 0.002.

ensure the reliability of this finding, we replicated the results in the written modality and additionally showed a general tendency for the cost to increase as the target-context similarity increased (manipulation of degree of similarity). The second experiment also replicated the contextual similarity cost. Furthermore, in neither experiment did the contextual similarity cost for individual participants correlate reliably between semantically- and segmentally-related pairs, compatible with selection control mechanisms that are specific to each production stage.

The contextual similarity costs in cyclic naming paradigms reflect a delicate balance between facilitatory effects that occur when similar words activate each other and the interference between such words, which is a negative side of such activation. This interference may arise because words with similar features actively compete during selection (e.g., Howard et al., 2006), but could also arise from the learning dynamics in a system without direct competition during selection (Navarrete, Del Prato, Peressotti, & Mahon, 2014; Oppenheim et al., 2010). In the latter, words that are activated through similar items but are not the targets of production undergo negative learning, which makes their subsequent retrieval more difficult. While past studies have separately shown the cost of contextual similarity in semantically- (see Schnur et al., 2014 and the references therein) and segmentally-related contexts (Breining et al., 2015), this is the first demonstration of the reliable interference induced by both contexts within individuals. We show that even though such interference is robust at both stages of production, the magnitude of the effect is not reliably correlated between these two stages in speakers. The stage-specificity of selection control in lexical retrieval is aligned with findings that point to the functional dissociation of semantic-lexical and lexical-phonological parts of the lexical retrieval system (e.g., Dell, Schwartz, Nozari, Faseyitan, & Branch Coslett, 2013), as well as neural evidence for the dissociation of control regions in semantic and phonological processing in comprehension (e.g., Bookheimer, 2002; Fiez, 1997; Gabrieli, Poldrack, & Desmond, 1998; Gough, Nobre, and Devlin, 2005).

The second goal of the study was to examine post-monitoring control and its relationship to selection control. Two models were considered. Model 1 specified a single locus for post-monitoring control at the level of segmentally-encoded output. Model 2 posited two independent post-monitoring control loci, one at the level of lexical selection and one at the level of segmental selection. As such, Model 1 predicted no interaction between the cost of reversal for related and unrelated words, and a positive correlation between reversal cost for semantically- and segmentally-related pairs. Model 2, on the other hand, predicted a reliable interaction and no correlation. The findings are consistent with the predictions of Model 1 and each is discussed below.

The first prediction of Model 1 was supported by an absence of an interaction between reversal costs and contextual similarity. Reversal costs were of comparable magnitude for words produced in similar and mixed contexts. This result is in agreement with results of Hartsuiker et al. (2005) who reported no influence of contextual overlap between the to-be-suppressed word and repair when the two were semantically related, or when they were segmentally related and the word was suppressed before making its way to overt production (see also PPI; e.g., Navarrete & Costa, 2005). Tydgat et al. (2012) replicated this finding for

phonological overlap between the two words in covert repairs, but found facilitation as a function of semantic overlap. In our findings, however, there was no evidence of facilitation induced by similar context when words were to be reversed. What might be the source of the different findings of these studies? Note that Tydgat et al.'s (2010) studies used an interruption paradigm, in which speakers had to abandon naming one picture as quickly as possible when a second picture appeared, instead naming the new object. Before the second picture appeared, no information about its identity was available to participants. In such cases, the first picture activates some of the features of the second picture, hence facilitating its production as in prime-target studies (e.g., Costa, Alario, & Caramazza, 2005; Mahon, Costa, Peterson, Vargas, & Caramazza, 2007; Wheeldon & Monsell, 1994). It is quite reasonable to expect production of a new and unexpected target to benefit from partial activation of its semantic features through semantic priming. For example, when the target is *cat*, activation of its semantic features leads to partial activation of words referring to other animals in so far as they share certain features with cat. Thus, if the next target is *dog*, it can benefit from the partial activation of its features by the previous target, compared to when it is followed by an unrelated item which shares no semantic features with it. However, insufficient activation of semantic features is rarely the problem in speech errors in neurotypical individuals. Even if the speaker has transient trouble with lexical retrieval, semantics of the target word are typically highly activated, hence the much more frequent occurrence of semantically-related errors compared to unrelated errors in neurologically intact speakers (Dell, Nozari, & Oppenheim, 2014). This feature was captured in the current design; both objects were known to the speaker, reducing the possible benefit of semantically-related items in activating a previously-unknown target.

Although the absence of a reliable interaction between contextual and reversal costs supports Model 1, an unreliable interaction may represent the lack of statistical power to find a significant effect. For this reason, it was critical to test the second prediction of Model 1, namely a reliable correlation between the reversal costs between semantically- and segmentally-related pairs. While neither experiment found a reliable correlation between selection control for the semantically- and segmentally-related pairs, the correlation between reversal costs for the two pair types was positive and reliable. Collectively, the absence of the interaction between the two costs, and the positive correlation between the costs for semantically- and segmentally-related pairs support Model 1, a model in which selection control is enforced at each selection level, but monitoring control is enforced only at the outcome of both selection processes.

While we believe that the design of Experiment 2 is one step closer to capturing the essence of monitoring and repair processes in word production, the design is not without limitations. One might object that in our reversal task, participants knew that they had to suppress a potent response on each trial (whereas such information is not available a priori on error trials in everyday speech). As such, our participants may have been overly prepared to suppress the dispreferred word. Note, however, that better preparation should, if anything, predict suppression at an earlier stage (e.g., during lexical selection as predicted by Model 2), whereas the results support a late focus for the deployment of post-monitoring control, with an unchanged cost of either semantic or segmental similarity cost in the reversed phase. Ultimately, we acknowledge that studying the details of monitoring and repair processes in

word production is methodologically challenging. More natural methods (e.g., spontaneous monitoring during picture naming) can be used in individuals with brain damage (e.g., Nozari et al., 2011; Riès, Xie, Haaland, Dronkers, & Knight, 2013), but are unlikely to yield enough errors in neurotypical individuals. As such, we are left with the option of using experimental paradigms to capture the cognitive processes underlying monitoring and repair as closely as possible, which we believe this study has accomplished in a reasonable way.

Finally, from a clinical perspective, the paradigm used, and the specific control processes tested in this study, are important for aphasia rehabilitation. Naming a small set of pictures several times in a cycle is a critical part of aphasia treatment, and treatment cycles often comprise pictures that are related in some way. Our results show that interference is to be expected in such conditions, and might be exaggerated in individuals with specific control deficits (e.g., Biegler, Crowther, & Martin, 2008). As such, training selection control abilities may help such individuals during treatment. Post-monitoring control is also critical, as individuals must suppress the often-incorrect response generated by the noisy, damaged production system in order to produce the correct response, a process that may also benefit from training. The results of our analyses of individual differences suggest that the best approach to training cognitive control in order to improve word production might be to separate training of selection control at the levels of lexical and segmental selection. However, post-monitoring control can be trained regardless of contextual similarity, and improvement in stopping the incorrect response can be expected regardless of the relationship between the error and the potential repair.

Conclusion

We studied two types of control processes needed for selection and repair during word production: selection control and post-monitoring control. Selection control was stage-dependent, varying depending on whether contextual similarity affected lexical or segmental selection. Post-monitoring control suppressed the dispreferred words, but did not modulate selection control. Furthermore, it was common to both stages of production. These findings suggest a model in which post-monitoring control operates after both lexical and segmental selection are completed.

Acknowledgments

References

Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language. 2013; 68(3):255–278. doi: 10.1016/j.jml. 2012.11.001.

Belke E, Meyer AS. Single and multiple object naming in healthy ageing. Language and Cognitive Processes. 2007; 22(8):1178–1211. doi: 10.1080/01690960701461541.

Belke E, Meyer AS, Damian MF. Refractory effects in picture naming as assessed in a semantic blocking paradigm. The Quarterly Journal of Experimental Psychology. 2005; 58(4):667–692. doi: 10.1080/02724980443000142. [PubMed: 16104101]

Biegler KA, Crowther JE, Martin RC. Consequences of an inhibition deficit for word production and comprehension: Evidence from the semantic blocking paradigm. Cognitive Neuropsychology. 2008; 25(4):493–527. doi: 10.1080/02643290701862316. [PubMed: 19086200]

Blackmer ER, Mitton JL. Theories of monitoring and the timing of repairs in spontaneous speech. Cognition. 1991; 39(3):173–194. [PubMed: 1841032]

Bloem I, van den Boogaard S, La Heij W. Semantic facilitation and semantic interference in language production: Further evidence for the conceptual selection model of lexical access. Journal of Memory and Language. 2004; 51(2):307–323. doi: 10.1016/j.jml.2004.05.001.

Bonin P, Peereman R, Fayol M. Do phonological codes constrain the selection of orthographic codes in written picture naming? Journal of Memory and Language. 2001; 45(4):688–720. doi: 10.1006/jmla.2000.2786.

Bookheimer S. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annual Review of Neuroscience. 2002; 25(1):151–188.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. Conflict monitoring and cognitive control. Psychological Review. 2001; 108(3):624. doi: 10.1037/0033-295X.108.3.624. [PubMed: 11488380]

Breining B, Nozari N, Rapp B. Does segmental overlap help or hurt? Evidence from blocked cyclic naming in spoken and written production. Psychonomic Bulletin & Review. 2015:1–7. doi: 10.3758/s13423-015-0900-x. [PubMed: 24847901]

Coltheart M, Woollams A, Kinoshita S, Perry C. A position-sensitive Stroop effect: Further evidence for a left-to-right component in print-to-speech conversion. Psychonomic Bulletin & Review. 1999; 6(3):456–463. doi: 10.3758/BF03210835. [PubMed: 12198784]

Costa A, Alario F-X, Caramazza A. On the categorical nature of the semantic interference effect in the picture-word interference paradigm. Psychonomic Bulletin & Review. 2005; 12(1):125–131. doi: 10.3758/BF03196357. [PubMed: 15948287]

Crowther JE, Martin RC. Lexical selection in the semantically blocked cyclic naming task: the role of cognitive control and learning. Frontiers in Human Neuroscience. 2014; 8(9) doi: 10.3389/fnhum.2014.00009.

Damian MF. Articulatory duration in single-word speech production. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2003; 29(3):416. doi: 10.1037/0278-7393.29.3.416.

Damian MF, Bowers JS. Effects of orthography on speech production in a form-preparation paradigm. Journal of Memory and Language. 2003; 49(1):119–132. doi: 10.1016/S0749-596X(03)00008-1.

Damian MF, Vigliocco G, Levelt WJ. Effects of semantic context in the naming of pictures and words. Cognition. 2001; 81(3):B77–B86. doi: 10.1016/S0010-0277(01)00135-4. [PubMed: 11483172]

Dell, GS., Nozari, N., Oppenheim, GM. Lexical access: Behavioral and computational considerations. In: Ferreira, V.Goldrick, M., Miozzo, M., editors. The Oxford Handbook of Language Production. Oxford University Press; 2014. p. 88-104.

Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA. Lexical access in aphasic and nonaphasic speakers. Psychological Review. 1997; 104(4):801. doi: 10.1037/0033-295X.104.4.801. [PubMed: 9337631]

Dell GS, Schwartz MF, Nozari N, Faseyitan O, Branch Coslett H. Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. Cognition. 2013; 128(3):380–396. doi: 10.1016/j.cognition.2013.05.007. [PubMed: 23765000]

Fiez JA. Phonology, semantics, and the role of the left inferior prefrontal cortex. Human Brain Mapping. 1997; 5(2):79–83. [PubMed: 10096412]

Gabrieli JD, Poldrack RA, Desmond JE. The role of left prefrontal cortex in language and memory. Proceedings of the National Academy of Sciences. 1998; 95(3):906–913. doi: 10.1073/pnas.95.3.906.

Goldrick M, Folk JR, Rapp B. Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. Journal of Memory and Language. 2010; 62(2):113–134. doi: 10.1016/j.jml.2009.11.008. [PubMed: 20161591]

Hartsuiker RJ, Pickering MJ, Jong NH. Semantic and phonological context effects in speech error repair. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2005; 31(5):921. doi: 10.1037/0278-7393.31.5.921.

Hickok G. Computational neuroanatomy of speech production. Nature Reviews Neuroscience. 2012; 13(2):135–145. doi: 10.1038/nrn3158.

Howard D, Nickels L, Coltheart M, Cole-Virtue J. Cumulative semantic inhibition in picture naming: Experimental and computational studies. Cognition. 2006; 100(3):464–482. doi: 10.1016/j.cognition.2005.02.006. [PubMed: 16413014]

January D, Trueswell JC, Thompson-Schill SL. Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. Journal of Cognitive Neuroscience. 2009; 21(12):2434–2444. [PubMed: 19199402]

Jescheniak JD, Oppermann F, Hantsch A, Wagner V, Mädebach A, Schriefers H. Do perceived context pictures automatically activate their phonological code? Experimental Psychology. 2009; 56(1):56. doi: 10.1027/1618-3169.56.1.56. [PubMed: 19261579]

Kan IP, Thompson-Schill SL. Effect of name agreement on prefrontal activity during overt and covert picture naming. Cognitive, Affective, & Behavioral Neuroscience. 2004; 4(1):43–57. doi: 10.3758/CABN.4.1.43.

Kroll JF, Stewart E. Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. Journal of Memory and Language. 1994; 33(2):149–174. doi: 10.1006/jmla.1994.1008.

Levelt WJ. Monitoring and self-repair in speech. Cognition. 1983; 14(1):41–104. doi: 10.1016/0010-0277(83)90026-4. [PubMed: 6685011]

Levelt, WJ. Speaking: From intention to articulation. MIT Press; Cambridge, MA: 1989.

Levelt WJ, Roelofs A, Meyer AS. A theory of lexical access in speech production. Behavioral and Brain Sciences. 1999; 22(01):1–38. doi: 10.1017/S0140525X99001776. [PubMed: 11301520]

Mahon BZ, Costa A, Peterson R, Vargas KA, Caramazza A. Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2007; 33(3):503. doi: 10.1037/0278-7393.33.3.503.

Meyer AS. The time course of phonological encoding in language production: The encoding of successive syllables of a word. Journal of Memory and Language. 1990; 29(5):524–545. doi: 10.1016/0749-596X(90)90050-A.

Meyer AS, Damian MF. Activation of distractor names in the picture-picture interference paradigm. Memory & Cognition. 2007; 35(3):494–503. doi: 10.3758/BF03193289. [PubMed: 17691148]

Miceli G, Benvegnu B, Capasso R, Caramazza A. The independence of phonological and orthographic lexical forms: Evidence from aphasia. Cognitive Neuropsychology. 1997; 14(1):35–69. doi: 10.1080/026432997381619.

Navarrete E, Costa A. Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. Journal of Memory and Language. 2005; 53(3):359–377. doi: 10.1016/j.jml.2005.05.001.

Navarrete E, Del Prato P, Peressotti F, Mahon BZ. Lexical selection is not by competition: Evidence from the blocked naming paradigm. Journal of Memory and Language. 2014; 76:253–272. doi: 10.1016/j.jml.2014.05.003. [PubMed: 25284954]

Nozari N, Dell GS, Schwartz MF. Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. Cognitive Psychology. 2011; 63(1):1–33. doi: 10.1016/j.cogpsych.2011.05.001. [PubMed: 21652015]

Nozari N, Thompson-Schill SL. More attention when speaking: Does it help or does it hurt? Neuropsychologia. 2013; 51(13):2770–2780. doi: 10.1016/j.neuropsychologia.2013.08.019. [PubMed: 24012690]

Nozari, N., Thompson-Schill, SL. Left Ventrolateral Prefrontal Cortex in Processing of Words and Sentences. In: Hickok, G., Small, SL., editors. The Neurobiology of Language. Academic Press; Waltham, MA: 2015. p. 569-588.

Oppenheim GM, Dell GS, Schwartz MF. The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. Cognition. 2010; 114(2):227–252. doi: 10.1016/j.cognition.2009.09.007. [PubMed: 19854436]

Oppermann F, Jescheniak JD, Görges F. Resolving competition when naming an object in a multiple-object display. Psychonomic Bulletin & Review. 2014; 21(1):78–84. doi: 10.3758/s13423-013-0465-5. [PubMed: 23761213]

O'Séaghdha PG, Frazer AK. The exception does not rule: Attention constrains form preparation in word production. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2014; 40(3):797. doi: 10.1037/a0035576.

Pickering MJ, Garrod S. An integrated theory of language production and comprehension. Behavioral and Brain Sciences. 2013; 36(04):329–347. doi: 10.1017/S0140525X12001495. [PubMed: 23789620]

Postma A. Detection of errors during speech production: A review of speech monitoring models. Cognition. 2000; 77(2):97–132. doi: 10.1016/S0010-0277(00)00090-1. [PubMed: 10986364]

Rapp B, Benzing L, Caramazza A. The autonomy of lexical orthography. Cognitive Neuropsychology. 1997; 14(1):71–104. doi: 10.1080/026432997381628.

Rapp B, Fischer-Baum S, Miozzo M. Modality and Morphology What We Write May Not Be What We Say. Psychological Science. 2015 0956797615573520. doi; 10.1177/0956797615573520.

Rastle K, Brysbaert M. Masked phonological priming effects in English: Are they real? Do they matter? Cognitive Psychology. 2006; 53(2):97–145. doi: 10.1016/j.cogpsych.2006.01.002. [PubMed: 16554045]

Redick TS, Shipstead Z, Harrison TL, Hicks KL, Fried DE, Hambrick DZ, Engle RW. No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. Journal of Experimental Psychology: General. 2013; 142(2):359. doi: 10.1037/a0029082. [PubMed: 22708717]

Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S. The role of the medial frontal cortex in cognitive control. Science. 2004; 306(5695):443–447. doi: 10.1126/science.1100301. [PubMed: 15486290]

Riès SK, Xie K, Haaland KY, Dronkers NF, Knight RT. Role of the lateral prefrontal cortex in speech monitoring. Frontiers in Human Neuroscience. 2013; 7(703) doi: 10.3389/fnhum.2013.00703.

Roelofs A. Phonological segments and features as planning units in speech production. Language and Cognitive Processes. 1999; 14(2):173–200. doi: 10.1080/016909699386338.

Roelofs A. Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. Psychological Review. 2003; 110(1):88. doi: 10.1037/0033-295X.110.1.88. [PubMed: 12529058]

Roelofs A. Tracing attention and the activation flow in spoken word planning using eye movements. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2008; 34(2):353. doi: 10.1037/0278-7393.34.2.353.

Sadat J, Martin CD, Costa A, Alario F-X. Reconciling phonological neighborhood effects in speech production through single trial analysis. Cognitive Psychology. 2014; 68:33–58. doi: 10.1016/j.cogpsych.2013.10.001. [PubMed: 24291531]

Schnur TT. The persistence of cumulative semantic interference during naming. Journal of Memory and Language. 2014; 75:27–44. doi: 10.1016/j.jml.2014.04.006.

Schnur TT, Schwartz MF, Brecher A, Hodgson C. Semantic interference during blocked-cyclic naming: Evidence from aphasia. Journal of Memory and Language. 2006; 54(2):199–227. doi: 10.1016/j.jml.2005.10.002.

Schnur TT, Schwartz MF, Kimberg DY, Hirshorn E, Coslett HB, Thompson-Schill SL. Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca's area. Proceedings of the National Academy of Sciences. 2009; 106(1):322–327. doi: 10.1073/pnas.0805874106.

Schriefers H, Meyer AS, Levelt WJ. Exploring the time course of lexical access in language production: Picture-word interference studies. Journal of Memory and Language. 1990; 29(1):86–102. doi: 10.1016/0749-596X(90)90011-N.

Shen XR, Damian MF, Stadthagen-Gonzalez H. Abstract graphemic representations support preparation of handwritten responses. Journal of Memory and Language. 2013; 68(2):69–84. doi: 10.1016/j.jml.2012.10.003.

Stroop JR. Studies of interference in serial verbal reactions. Journal of Experimental Psychology. 1935; 18(6):643. doi: 10.1037/0096-3445.121.1.15.

Szekely A, Jacobsen T, D'Amico S, Devescovi A, Andonova E, Herron D, Wicha N. A new on-line resource for psycholinguistic studies. Journal of Memory and Language. 2004; 51(2):247–250. doi: 10.1016/j.jml.2004.03.002. [PubMed: 23002322]

Thompson-Schill SL, Swick D, Farah MJ, D'Esposito M, Kan IP, Knight RT. Verb generation in patients with focal frontal lesions: A neuropsychological test of neuroimaging findings. Proceedings of the National Academy of Sciences. 1998; 95(26):15855–15860. doi: 10.1073/pnas. 95.26.15855.

Tydgat I, Diependaele K, Hartsuiker RJ, Pickering MJ. How lingering representations of abandoned context words affect speech production. Acta Psychologica. 2012; 140(3):218–229. doi: 10.1016/j.actpsy.2012.02.004. [PubMed: 22673067]

Verbruggen F, Logan GD. Automatic and controlled response inhibition: associative learning in the go/no-go and stop-signal paradigms. Journal of Experimental Psychology: General. 2008; 137(4): 649. doi: 10.1037/a0013170. [PubMed: 18999358]

Wheeldon LR, Monsell S. Inhibition of spoken word production by priming a semantic competitor. Journal of Memory and Language. 1994; 33(3):332–356. doi: 10.1006/jmla.1994.1016.
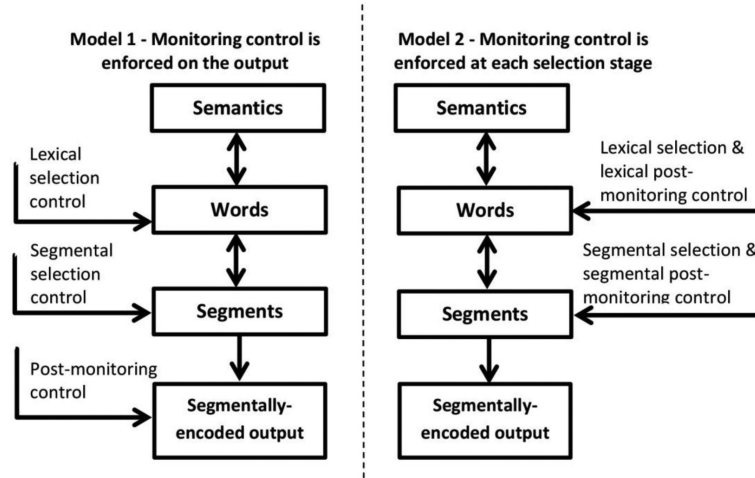
Figure 1.
Two possible models of the interaction between selection and post-monitoring control. In model 1, post-monitoring control is independent of selection control, and is applied to the segmentally-encoded output. In model 2, the post-monitoring control is applied to the same level as selection control, which differs for semantically- and segmentally-related pairs.
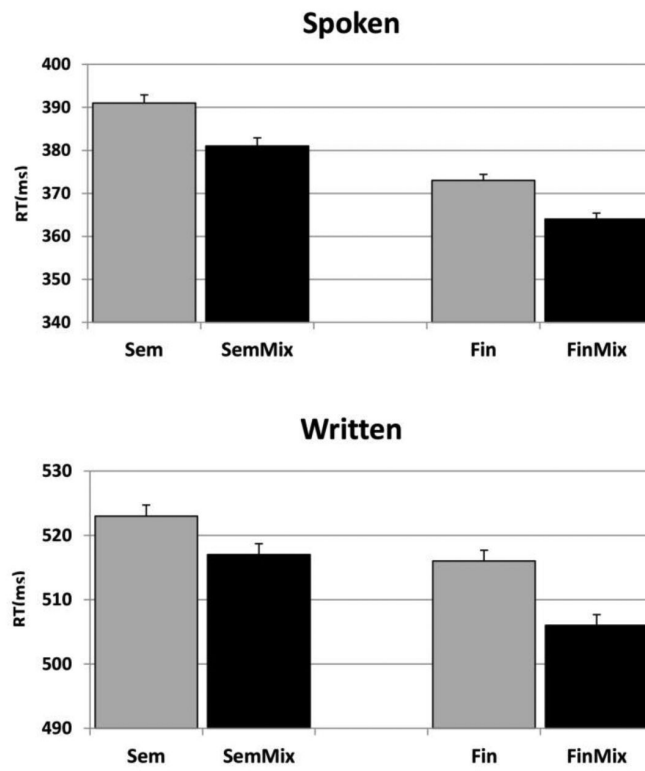
Figure 2.
RTs (±SE) for related and mixed conditions in spoken (upper panel) and written (lower panel) modalities. Fin = Final segmental overlap; Sem = semantic. The figure represents means of subject means.
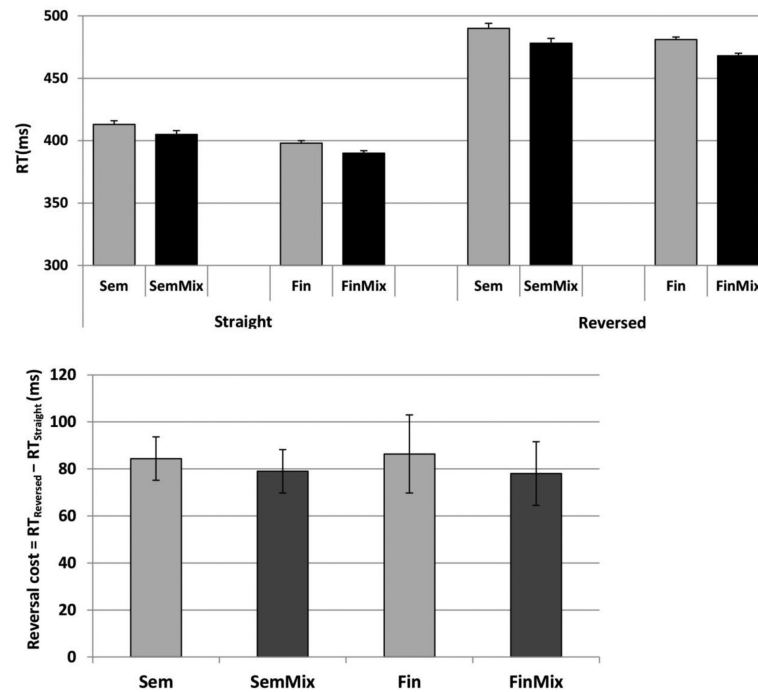
Figure 3.
Results of Experiment 2. Upper panel shows RT (+SE) in semantic and final segment
overlap conditions and their corresponding mixed conditions, in straight and reversed
phases. It can be seen that the contextual cost is not reduced in the reversed phase. The lower
panel graphs the reversal cost (± 95%CI) in each condition. This panel shows that reversal
imposes a cost that is close in magnitude between related and mixed conditions. Sem =
Semantic; Fin = Final segmental overlap. Both figures represent means of subject means.
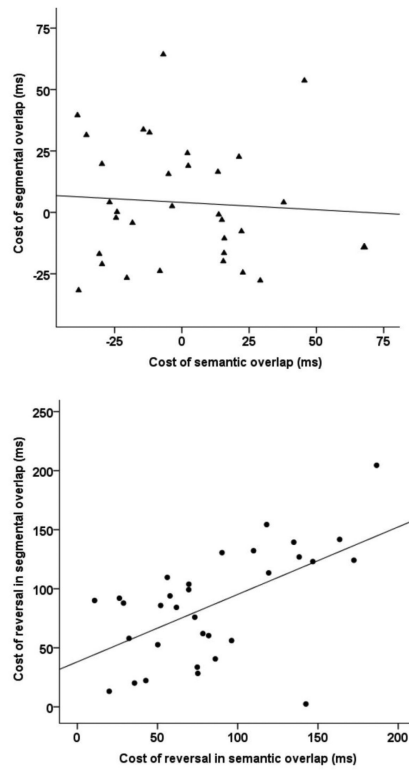
Figure 4.
Correlation between the contextual similarity cost (upper panel) and reversal cost (lower panel) for pairs that were related semantically, or in their final segments.

Table 1

Error proportions in Semantic and Final overlap conditions and their corresponding control conditions, for spoken and written modalities in Experiment 1.

Spoken	Semantic	Final
Related	0.06	0.03
Mixed	0.05	0.04
Written		
Related	0.007	0.007
Mixed	0.008	0.004

Table 2

RTs in Semantic and Final overlap conditions (±SE) for high and low similarity conditions in spoken and written modalities in Experiment 1. SE's are shown as a measure of variability, and include both within-subject and between-subject sources of variance, only the former of which is relevant for comparisons.

Spoken		Interference
SemHi	398(±11 ms)	18 ms
MixedSemHi	380(±10 ms)	
SemLo	387(±11 ms)	5 ms
MixedSemLo	382(±9 ms)	
FinHi	376(±10 ms)	15 ms
MixedFinHi	361(±8 ms)	
FinLo	371(±8 ms)	4 ms
MixedFinLo	367(±7 ms)	
Written		
SemHi	524(±1 5 ms)	10 ms
MixedSemHi	514(±17 ms)	
SemLo	524(±13 ms)	4 ms
MixedSemLo	520(±16 ms)	
FinHi	515(±18 ms)	12 ms
MixedFinHi	503(±19 ms)	
FinLo	518(±17 ms)	9 ms
MixedFinLo	509(±18 ms)	

Table 3

Results of the analysis of the semantic model in Exp. 1

Fixed effects	Coefficient	SE	t	P
Intercept	6.13662	0.034936	175.66	<0.001
Context (related/mixed)	0.015021	0.006597	2.28	0.03199
Modality (written/spoken)	0.300758	0.028103	10.7	<0.001
Degree of similarity (high/low)	0.011109	0.012262	0.91	0.37241
Order (spoken first)	−0.135073	0.044996	−3	0.00636
Context * Modality	−0.005336	0.008066	−0.66	0.50828
Context * Degree of similarity	−0.032901	0.008065	−4.08	<0.001
Modality * Degree of similarity	0.011091	0.011375	0.97	0.32958
Context * Modality * Degree of similarity	0.023113	0.016129	1.43	0.15188

Random effects	Variance
Subject intercept	0.013947
Context|Subject slope	0.0006813
Modality|Subject slope	0.0189361
Degree of similarity|Subject slope	0.0009699
Item intercept	0.0003168
Residual	0.0489702

Table 4

Results of the analysis of the segmental model in Exp. 1

Fixed effects	Coefficient	SE	t	P
Intercept	6.100194	0.035037	174.11	<0.001
Context (related/mixed)	0.02076	0.004195	4.95	<0.001
Modality (written/spoken)	0.32288	0.025304	12.76	<0.001
Degree of similarity (high/low)	0.014495	0.00949	1.53	0.1394
Order (spoken first)	−0.127578	0.045154	−2.83	0.0096
Context * Modality	−0.001916	0.007808	−0.25	0.8061
Context * Degree of similarity	−0.017961	0.007807	−2.3	0.0214
Modality * Degree of similarity	−0.006609	0.011039	−0.6	0.5494
Context * Modality * Degree of similarity	0.027823	0.015615	1.78	0.0748

Random effects	Variance
Subject intercept	0.0142
Context|Subject slope	0.0000589
Modality|Subject slope	0.0153
Degree of similarity|Subject slope	0.000225
Item intercept	0.0002
Residual	0.0462

Table 5

Results of the data analysis in Exp. 2

Fixed effects	Coefficient	SE	t	p
Intercept	5.9912147	0.0163621	366.2	<0.001
Context (related/mixed)	0.0151816	0.0073924	2.1	0.0407
Phase (straight/reversed)	−0.1688408	0.0121218	−13.9	<0.001
Relation (segmental/semantic)	0.024256	0.0112449	2.2	0.0382
Context * Phase	0.0007388	0.0143059	0.1	0.9588
Context * Relation	−0.0085109	0.0101345	−0.8	0.401
Phase * Relation	0.0068794	0.0101352	0.7	0.4973
Context * Phase * Relation	−0.0258981	0.0202682	−1.3	0.2014

Random effects	Variance			
Subject intercept	0.0067617			
Contex|Subject slope	0.0001112			
Phase|Subject slope	0.0030608			
Relation|Subject slope	0.0004343			
Item intercept	0.02084			
Residual	0.0928654			