

1
2
3 **Is repairing speech errors an automatic or a controlled process?**
4
5

6 **Insights from the relationship between error and repair probabilities in English and Spanish**
7

8
9 Nazbanou Nozari^{a,b*}, Clara D. Martin^{c,d}, and Nicholas McCloskey^e
10

11
12 ^aDepartment of Neurology, Johns Hopkins University, Baltimore, USA; ^bDepartment of Cognitive Science,
13 Johns Hopkins University, Baltimore, USA; ^cBasque Center on Cognition, Brain and Language, San
14 Sebastian, Spain; ^dIkerbasque - Basque Foundation for Science, Bilbao, Spain; ^eCenter for Substance
15 Abuse Research, Lewis Katz School of Medicine, Temple University, Philadelphia, USA.
16
17
18
19
20
21

22 Correspondences should be addressed to:
23

24 Nazbanou Nozari, MD, PhD
25

26 Address: 1629 Thames Street, Suite 350, Baltimore, MD 21231, USA
27

28 Phone: 443-287-1712
29

30 Fax: 410-955-0188
31

32 E-mail: nozari@jhu.edu
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

Speakers can correct their speech errors, but the mechanisms behind repairs are still unclear. Some findings, such as the speed of repairs and speakers' occasional unawareness of them, point to an automatic repair process. This paper reports a finding that challenges a purely automatic repair process. Specifically, we show that as error rate increases, so does the proportion of repairs. Twenty highly-proficient English-Spanish bilinguals described dynamic visual events in real time (e.g., "*The blue bottle disappears behind the brown curtain*") in English and Spanish blocks. Both error rates and proportion of corrected errors were higher on (a) noun phrase (NP)2 vs. NP1, and (b) word1 (adjective in English and noun in Spanish) vs. word2 within the NP. These results show a consistent relationship between error and repair probabilities, disentangled from position, compatible with a model in which greater control is recruited in error-prone situations to enhance the effectiveness of repair.

Keywords: speech errors, monitoring, repair, cognitive control, bilingualism, sentence production, semantic interference, incremental planning

Introduction

Speakers detect and correct their speech errors from an early age. Children show evidence of spontaneous error correction almost as soon as they start producing speech (Karmiloff-Smith, 1986), and this ability increases steadily with age (Hanley, Cortis, Budd, & Nozari, 2016) until adulthood. While the mechanisms of error *detection* have received much attention in the literature (see Nozari & Novick, 2017 for a review), relatively little is known about the mechanisms underlying error *correction*. This is in part due to the sparsity of empirical reports on the properties of repairs, which are necessary for proposing a formal model of repairing errors in speech production. This paper investigates one of the key properties of repairs, namely whether they are products of an automatic or a controlled process.

Repair as an automatic process

The question of automaticity is one of the oldest and most debated issues in psychology, dating back to William James in the 19th century (James, 1890). Various criteria, as well as different approaches such as feature-based vs. construct-based views, have been proposed to define automatic processes, and even feature-based approaches have created disagreement on whether judgments of automaticity should hinge on all or a subset of critical features, giving rise to all-or-none vs. decompositional views, respectively (see Moors & De Houwer, 2006 for a comprehensive review). Part of this divergence is due to the large diversity in the scopes and domains to which applying definitions of automaticity has been attempted: for example, automaticity has been investigated in the context of tasks as simple as retrieving the name of a picture to much more complex tasks such as driving from home to work in a crowded city. Despite the differences, however, there is at least some agreement that certain criteria are highly relevant to the question of automaticity, and have also been useful in evaluating automaticity in the language processing system. Some of these criteria include being *unintentional*, *goal-independent*,

1
2
3 *stimulus-driven, fast, efficient (i.e., effortless), and unconscious.* As pointed out by Moors and De
4
5 Houwer (2006), even partial presence of these features is useful for placing a task closer to the
6
7 automatic end of the automatic vs. controlled spectrum.
8

9
10 In language processing, Fodor's seminal essay on "The modularity of mind" presented a strong
11
12 defense of automaticity in language comprehension by arguing for its fast, efficient, and mostly
13
14 unintentional and unconscious nature (as in overhearing others without planning to do so; Fodor, 1983).
15
16 Interestingly, there were only occasional references to language *production* in that essay, and those few
17
18 remarks indicated that Fodor viewed language production to be fundamentally different from
19
20 comprehension in that the former stemmed from thought. The link to thought processes, in Fodor's
21
22 view, puts language production in the domain of conscious planning, as opposed to automatic
23
24 processing which he held responsible for comprehension, although he briefly mentioned that the motor
25
26 part of production may indeed behave more similarly to comprehension in terms of automaticity (Fodor,
27
28 1983, p. 42). Fodor's dilemma with language production reflects a larger problem of evaluating the
29
30 automaticity of production in light of the standard criteria for automaticity. The most important caveat
31
32 is that language production is, on the one hand, a clearly intentional and goal-oriented task, and such
33
34 goals affect the inner processes of production such as the placement of a selection criterion (Nozari &
35
36 Hepner, 2018). On the other hand, many production processes that are carried out to map meaning
37
38 onto sound, such as spreading activation, do meet many criteria of automaticity such as being fast,
39
40 efficient and largely subconscious (Dell, 1986). Still some aspects of the same operations, such as
41
42 resolution of conflict between competing alternatives during both lexical and syntactic processing, seem
43
44 to require control (see Hartsuiker & Moors, 2016 and Nozari, 2018 for comprehensive reviews).
45
46
47
48
49

50 In the current context, we are less concerned with an ontological classification of repairs as automatic
51
52 or controlled. The goal, instead, is to define properties of repairs that would bring us closer to
53
54 developing a functional model of error repairs in speech production. To this end, we will employ some
55
56
57
58
59
60

1
2
3 of the criteria previously proposed to distinguish automatic vs. controlled processes, but we are not
4
5 committing to any specific theory of automaticity.
6

7
8 In the non-linguistic domain, fast, efficient and sometimes unconscious repairs have been reported in
9
10 tasks such as the anti-saccade task. In such tasks participants must suppress the natural urge to saccade
11
12 towards a pop-up stimulus appearing on one side of the visual field and instead move their eyes to the
13
14 opposite side. Several studies have shown that participants sometimes correct a saccade (e.g., move
15
16 their eyes to the left, if they had originally performed a saccade to the right), without awareness of
17
18 having committed and corrected an error (Endrass, Reuter, & Kathmann, 2007; Nieuwenhuis,
19
20 Ridderinkhof, Blom, Band, & Kok, 2001; Wessel, Danielmeier, & Ullsperger, 2011).
21
22

23
24 Although not directly speaking to the repair process per se, findings on post-error adjustments also
25
26 imply that these processes do not depend heavily on conscious processing. One such adjustment is post-
27
28 error slowing (e.g., Notebaert et al., 2009): participants' tendency to respond more slowly on the trial
29
30 after an error trial, which has also been shown in the context of speech errors (Freund & Nozari, 2018).
31
32 In non-verbal tasks, post-error slowing has been observed on trials where participants were not
33
34 necessarily aware of the error (Hester, Foxe, Molholm, Shpaner, & Garavan, 2005). In a clever study,
35
36 Hester, Simoes-Franklin, and Garavan (2007) showed that cocaine users, despite having significantly
37
38 poorer error awareness than non-users, showed nevertheless similar post-slowing behavior.
39
40

41
42 These and similar pieces of evidence from the non-verbal tasks suggest that post-error adjustments,
43
44 including repairs, may be at least partially automatic. The tasks from which these findings were
45
46 recovered, however, are inarguably much simpler than language production, and often have a very
47
48 limited response set. It is thus important to consider whether an automatic model of repair is even
49
50 viable in language production.
51
52

53 54 **Evidence pointing to an automatic repair process in language production** 55 56

1
2
3 At the global task level (i.e., producing the correct word from meaning), it is clear that production has
4 an intention and a goal. On many occasions, there is no external “stimulus” triggering production, so the
5 production process per se cannot be defined as a purely stimulus-driven process. It is thus reasonable to
6
7
8
9
10 ask whether a repair process that does not meet the explicit criterion of intentionality is even viable. The
11
12 answer is that although the overall production process has a goal, one could envision a repair process
13
14 that is partially unintentional. An example of such a process would be simply selecting the next most
15
16 highly activated item once the produced item is deemed to be an error (see Nooteboom & Quené, 2019,
17
18 for a similar view). In such a case, the repair process did not set and follow a new goal. Instead, a simple
19
20 operation, i.e., selection, was triggered when an error was signaled. It is thus theoretically possible to
21
22 have a repair model that is at least partially unintentional and in some sense “stimulus-driven”, i.e.,
23
24 driven by a fixed prior event which in this case is an error signal. In light of this, one can examine
25
26 whether speech error repairs meet some of the other criteria for automaticity.
27
28
29

30 Repairing speech errors certainly meets the criteria of fast and efficient processing: errors can be
31
32 repaired very quickly with very little temporal gap between the point of error interruption and the start
33
34 of the repair (sometimes as short as 0 ms; Blackmer & Mitton, 1991; Hartsuiker & Kolk, 2001;
35
36 Nooteboom & Quené, 2019). While not incompatible with a very efficient controlled process, fast
37
38 operations are often taken to indicate some degree of automaticity (Fodor, 1983; Nozari, 2018).
39
40 Complementing these findings, the nature of repairs in certain individuals with post-stroke aphasia is
41
42 also relevant to this debate: some such individuals, despite having difficulty with primary production
43
44 processes, apply numerous fast repairs to their errors. In conduction aphasia this is called *conduite*
45
46 *d’approche* and is often observed in phonological errors (e.g., Target = igloo: response = /aj-, aj-, ajk-,
47
48 ajgpl, ajpg-, ajglu, ej, iglu, ajglu, rgglu, glu, o, ajglu, lɹJglu, li-, gli-, ajglu/, igloo, /iglu/, igloo; (Kohn, 1984)).
49
50 The interesting finding here is that repair attempts do not always bring the response closer to the target.
51
52
53
54 Sometimes the affected individuals even produce the correct response in the course of their multiple
55
56
57
58
59
60

1
2
3 attempts (the bold and underlined response in the example given here), but move right along to
4
5 producing other incorrect responses. Similar correction behavior is sometimes observed in non-
6
7 conduction aphasics attempting to correct their lexical errors. For example, Nozari (2019) reported an
8
9 individual with intact comprehension and predominantly semantic errors in production who frequently
10
11 attempted to correct his errors, but sometimes unwittingly passed over a correct repair (e.g., Target =
12
13 orange; response = apple, pineapple, pumpkin, orange, pineapple, peach?). This pattern is exactly what
14
15 is predicted by a repair process which automatically outputs the activated alternatives without tight
16
17 control over matching the response to the target.
18
19

20
21 Direct evidence for unconscious repair of speech errors is hard to obtain in speech production of
22
23 neurotypical adult speakers, because single word production (e.g., picture naming) rarely leads to errors
24
25 to begin with, and failed awareness of errors and repairs in multi-word utterances may reflect memory
26
27 lapses as opposed to genuine lack of awareness at the time of repair. Nevertheless, apart from
28
29 anecdotal evidence of unconscious repairs (Laver, 1973), there is some evidence that repairs without full
30
31 consciousness over the errors may be possible. For example, Postma and Noordanus (1996) asked
32
33 participants to press a button whenever they detected an error in their speech. They found that
34
35 occasionally self-repairs were unaccompanied by a button press, although this might simply mean that
36
37 the participant temporarily forgot the task instructions. In a just-completed study in our lab, participants
38
39 heard single words and typed them under time pressure without immediately seeing the results on the
40
41 screen. After each trial, they were asked whether they made a mistake. On 359 occasions, participants
42
43 denied having made an error, even though they had made an error and had repaired it. These reports
44
45 were not driven by the outcome (i.e., participants did not deny having made an error more often if the
46
47 final response was correct), and were observed in the majority of participants (80% out of 60). Still, this
48
49 effect may be exclusive to typing and not extendable to oral production. Finally, it has been reported
50
51
52
53
54
55
56
57
58
59
60

1
2
3 that children as young as 2 years of age, who lack awareness over what went wrong in their speech, can
4
5 nevertheless repair their errors (Clark, 1978; Karmiloff-Smith, 1986).
6

7 To summarize, although direct evidence for an automatic repair process is hard to obtain for language
8
9 production, various pieces of evidence point to the possibility of such a process.
10
11

12 13 14 **Evidence pointing to a controlled repair process in language production** 15

16 There is also indirect evidence for a potential influence of attentional and control processes on error
17
18 correction. For one thing, the percentage of corrected errors in speech is relatively low, which could
19
20 suggest that only errors that are attended to are corrected. For example, Nooteboom's (1980) analysis
21
22 of Meringer's speech corpus revealed a 75% correction rate for phonological errors and only a 53%
23
24 correction rate for semantic errors. Lapses in attention could explain missed repairs, but these may also
25
26 result from an automatic process that does not have a high hit rate. A more convincing piece of evidence
27
28 for the attentional account is that when instructed to correct their errors in the experimental setting,
29
30 speakers' repair rate often goes up, showing that they can strategically allocate more resources to
31
32 increase the repair rate. But even under these circumstances, the percentage of repaired errors often
33
34 remains below 80%, showing some form of resource limitation, generally aligned with the claims of
35
36 attentional effects on repair processing (Levelt, 1983; Oomen, Postma, & Kolk, 2005), although resource
37
38 limitation is not a unique feature of controlled processes (Moors & De Houwer, 2006). In a similar vein,
39
40 error detection performance in children increases with age, with 5, 6, and 8 year olds correcting on
41
42 average 40%, 50% and 70% of their errors in a sentence production task (Hanley et al., 2016). This is the
43
44 age when children's attentional and executive control abilities also develop fast (Weighall, 2008), but
45
46 the influence of other maturation processes cannot be ruled out.
47
48
49
50

51
52 The most convincing evidence for the influence of attention on repairs is the finding that the
53
54 percentage of detected errors increases towards the end of the utterance. Levelt (1983) had participants
55
56

1
2
3 describe colored objects in a network and investigated the proportion of corrected errors on colors that
4 were 0, 1, 2, and 3 syllables away from the end of the phrase (e.g., “And then you come to the *blue*” = 0
5 syllable gap; “There is a *yellow* node” = 1 syllable gap, etc.). He found that the correction rate for phrase-
6 final errors (i.e., gap 0) was 57%. In comparison, less than 20% of errors in longer gaps were corrected.
7
8 Levelt (1989) interpreted this finding as an influence of attention on monitoring: attentional resources
9 are used for primary production processes (e.g., message planning, selecting the lexical items, selecting
10 a syntactic frame, etc.) at earlier parts of the sentence, but are shifted to monitoring processes towards
11 the end of the phrase, when they are no longer needed for primary production processes. This
12 explanation is certainly plausible, but there is another striking finding there. The total number of errors
13 in the gap 0 position is 329, 4.5 times greater than the number of errors in gap 1 ($n = 73$).
14
15

16 The much greater rate of errors towards the end of the utterance suggests a three-way relationship
17 between position, error rate, and correction rate: Correction rate is highest at the end position, which
18 also happens to be the most error-prone position. There are thus two possibilities: a) that the higher
19 correction rate is directly related to position and indirectly related to error rates. This account maintains
20 that speakers strategically shift their attentional processes from primary production to monitoring
21 processes from the beginning to the end of the sentence, as the more primary production processes
22 have been completed towards the end of the sentence. According to this account, the higher error rate
23 in the final position is either irrelevant to repair rates, or may be a result of the strategic shift of
24 attentional processes towards monitoring and repair instead of primary production processes. Either
25 way, it does not view the higher probability of errors as the cause of the higher probability of repairs.
26
27 Empirically, this account predicts that later positions in the sentence should be associated with a higher
28 probability of repairs, as more primary production processes have been completed in later compared to
29 earlier positions in the sentence.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 b) The second possibility is that the repair rate is directly related to the error rate, and only
4
5 accidentally related to position. The greater proportion of repairs in more error-prone situations could
6
7 be explained as follows: when production becomes more difficult, the increased need for control leads
8
9 to the deployment of additional control resources (e.g., Freund & Nozari, 2018). If unsuccessful in
10
11 preventing errors from surfacing in overt speech, these newly deployed resources could at least catch
12
13 them after they are uttered. Therefore this account predicts that the probability of repairs may be
14
15 higher in earlier than later positions in the sentence, if those earlier positions are more error-prone.
16
17
18
19
20

21 **Current study**

22
23 While the characteristics of phonological repairs have been studied in some detail (Nootboom &
24
25 Quené, 2013a,b; 2015; 2017; 2019), much less attention has been paid to studying lexical repairs. There
26
27 are, however, good reasons for studying lexical repairs separately from segmental repairs, as the two
28
29 show different properties. For example, a study of spontaneous repairs in picture naming in aphasia
30
31 found that phonological repairs were started faster than lexical repairs, even after controlling for the
32
33 degree of phonological overlap in lexical and segmental error-repair pairs (Schuchard, Middleton, &
34
35 Schwartz, 2017). This difference might in turn point to at least partially different mechanisms for
36
37 detecting and repairing lexical and segmental errors. For example, forward models (e.g., Hickok, 2012),
38
39 which are highly plausible for monitoring segmental errors, are much less readily applicable to the
40
41 detection of lexical errors, while other mechanisms such as conflict detection can be easily applied to
42
43 lexical error detection/repair (Nozari & Novick, 2017).
44
45
46
47

48 One notable exception in the relatively understudied field of lexical repairs are studies in which
49
50 participants were externally prompted to change their response, e.g., because the picture they are
51
52 naming suddenly changed while they were producing its name (e.g., Hartsuiker, Catchpole, de Jong, &
53
54 Pickering, 2008; Hartsuiker, Pickering, & De Jong, 2005; Tydgat, Diependaele, Hartsuiker, & Pickering,
55
56
57
58
59
60

1
2
3 2012; Tydgat, Stevens, Hartsuiker, & Pickering, 2011; Van Wijk & Kempen, 1987). While valuable in
4
5 evaluating many aspects of the repair process including, but not limited to, potential interactions
6
7 between externally induced control processes and internal dynamics of the language production system
8
9 (Nozari, Freund, Breining, Rapp, & Gordon, 2016), such paradigms do not capture other aspects of the
10
11 monitoring and repair processes at work during everyday speech production. Most importantly, they
12
13 overlook a key feature of repairs, namely that the repair does not just spring into the speaker's
14
15 consciousness after an error is detected, but in many cases has been actively competing with the error
16
17 for some time before the error has surfaced. The fact that the repair is already highly activated in the
18
19 production system by the time of error detection is of vital importance for quick and subconscious
20
21 repairs. This, in turn, calls for studies that capture lexical repairs under conditions in which co-activation
22
23 of the reparandum and the repair is a natural consequence of planning an utterance from a conceptual
24
25 message (e.g., Levelt, 1983). The current study was designed with this goal in mind. Specifically, the
26
27 design aimed to establish whether the probability of lexical repairs fluctuated with changing control
28
29 demands and resources, as a function of error position in the sentence or error probability.
30
31
32
33

34 As discussed earlier, several pieces of evidence suggest some degree of automaticity in repairing
35
36 speech errors. Performance in pure automatic processes that are not subject to regulation by control
37
38 processes such as attention is usually stable within the same individual. For example, if the repair
39
40 process entails a mechanism of "replacing the rejected response with the next most highly activated
41
42 response", there is no reason that this mechanism should be implemented more frequently at the end
43
44 of the sentence than at its beginning. The fact that the proportion of repaired errors is different in
45
46 different positions (Levelt, 1983, 1989) points to a process that is at least to some degree subject to
47
48 fluctuations of control. To better understand the relationship between control and repairs, we must first
49
50 disentangle the relationship between position, error probability, and repair probability. To this end, we
51
52 conducted an experiment which allowed us to a) attempt to replicate Levelt's (1983) findings of the
53
54
55
56
57
58
59

1
2
3 three-way relationship between position, error and repair probability, and b) to investigate whether the
4 probability of repair follows position or the probability of errors, and in which direction.
5
6

7 To capture the cognitive processes involved in producing sentences from meaning, with the purpose
8 of a communicative message, we used a paradigm similar to Levelt's (1983). In the "Haunted Hotel"
9 paradigm, participants were presented with objects in a hotel room (e.g., window, curtain, suitcase,
10 telephone). They then learned that they would be viewing scenes of a haunted room, where these
11 objects moved around and interacted with one another. They were instructed to describe what was
12 going on in the room to a confederate (the experimenter) under time pressure. Each event consisted of
13 two objects performing one action, e.g., "*The blue bottle disappears behind the brown curtain*", and
14 could always be described using a [NP1 (Noun Phrase) + verb + NP2] structure. NPs always consisted of a
15 determiner, an adjective, and a noun. The choice of this type of event (and its corresponding structure)
16 allows us to conduct two sets of analyses.
17
18
19
20
21
22
23
24
25
26
27
28
29

30 Analysis 1 aimed to replicate the findings of Levelt (1983), by comparing repair rates on NP2 vs. NP1.
31 The NPs were semantically related in two ways: (a) colors which form a taxonomically related category,
32 and (b) nouns that are made to be semantically related in the experiment by creating the common
33 theme of "objects found in the haunted hotel room" which has been introduced to the participants in
34 the orientation phase. Abdel Rahman and Melinger (2011) showed that creating such ad hoc categories
35 is an effective way to induce semantic interference. Because the NPs are semantically related and are
36 named repeatedly, we expected significantly more errors on NP2 compared to NP1, because of the well-
37 established semantic blocking effect (Belke, Meyer, & Damian, 2005; Belke & Stielow, 2013; Damian &
38 Als, 2005; Damian & Bowers, 2003; Nozari et al., 2016; Schnur, Schwartz, Brecher, & Hodgson, 2006;
39 Schnur et al., 2009). This effect, which we suspect also underlies the significantly larger number of errors
40 in the final position in Levelt (1983), would predict higher error rates in the later compared to earlier
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 positions in the sentence. If the relationship between position, error rate, and probability of repairs
4
5 observed in Levelt (1983) is robust, analysis 1 should be able to replicate those findings.
6

7
8 Analysis 2 then aimed to disentangle the effect of position and probability of errors on the probability
9
10 of repairs, by comparing error rates on the two elements of the NP (color and noun). Findings from
11
12 similar paradigms used in previous experiments suggest that the time pressure of describing dynamic
13
14 events encourages incremental planning (Arnold & Nozari, 2017). Under these circumstances, there is
15
16 always less time to plan the earlier element of the NP, compared to the later element (which can be
17
18 planned as the earlier element is being articulated). For example, Nozari, Arnold, & Thompson-Schill
19
20 (2014) found that error rates were comparable on adjectives and nouns in NPs such as “red trapezoid”
21
22 even though the lexical frequency of the noun was much lower than the adjective, which may indicate
23
24 that the retrieval of the low-frequency noun could have benefitted from the additional planning time
25
26 provided during the articulation of the adjective. Similarly, anodal stimulation of the left prefrontal
27
28 cortex was less successful in decreasing the error rates on adjectives compared to nouns, pointing to the
29
30 limited planning time to produce the early element of the NP as a serious obstacle in producing the
31
32 correct target word. Based on these findings we predicted that when incremental planning is
33
34 encouraged by the task (e.g., describing dynamic events under time pressure), earlier elements of the
35
36 NP would be more error prone than later elements. Therefore, in a sentence like “*The blue bottle*
37
38 *disappears behind the brown curtain*”, we would expect more errors on “blue” and “brown” (*word1*)
39
40 compared to “bottle” and “curtain” (*word2*).
41
42
43
44

45
46 If position is the determining factor, the probability of repair should be higher on *word2*. If, on the
47
48 other hand, error rate is the critical factor, repair probability should be higher on *word1*. But there is a
49
50 problem here: in English adjectives come before nouns, thus the position within NP is systematically
51
52 confounded with part of speech. It is thus impossible to study the true effect of position on the
53
54 proportion of repairs unless we can disentangle position and part of speech. The solution to this
55
56
57
58
59
60

1
2
3 problem is to use a language in which the word order within NP is reversed. For example, the Spanish
4 equivalent of the sentence above would be “La botella azul desaparece por detrás de la cortina marrón.”
5
6 where “botella” and “cortina” are now word1, and “azul” and “marrón” are word2. If our assumption
7
8 about the higher probability of error on earlier NP elements holds regardless of part of speech, English
9
10 and Spanish should show very similar pattern of errors based on position. We can then test the effect of
11
12 position on repairs without the confound of part of speech.
13
14
15

16 Adding Spanish has two additional advantages: 1) it provides an opportunity for replication across
17 both analyses. 2) It allows us to examine whether the effects are language-specific or not. As far as
18
19 predictions go, there is no clear reason to expect that one account would hold in one language but not
20
21 the other, at least as long as speakers are more or less equally proficient in producing sentences in both
22
23 languages. To this end, we recruited 20 highly proficient Spanish-English bilingual speakers who still
24
25 used Spanish with at least some family members and friends, although they lived and worked in an
26
27 English-speaking environment (Baltimore, Maryland, United States). They completed four blocks of the
28
29 same task, two in English, and two in Spanish in counterbalanced order, generating a corpus of 500+
30
31 speech errors on NPs. Note that our primary interest in this experiment is not to compare monitoring in
32
33 L1 and L2. As far as proficiency and fluency are comparable for the task at hand, we would expect fairly
34
35 comparable baseline performance in terms of number of errors and repairs across the two languages.
36
37 Thus, in this case, the use of two languages is aimed at replicating the same findings in two languages,
38
39 and ruling out the confound of part of speech in order to have a clean test of the two hypotheses of
40
41 main interest.
42
43
44
45
46

47 Below is a summary of the three possible outcomes of the experiment and accounts supported by
48
49 each:
50

51
52 1) No relationship between error rates, position, and repair proportions (i.e., a failure to replicate
53
54 Levelt’s 1983 findings). This scenario would be most compatible with an account in which repair
55
56
57
58
59
60

1
2
3 processes are not sensitive to change in control demands and resources, i.e., a *purely automatic*
4
5 *account*.

6
7 2) An increase in repair proportions for later —as opposed to earlier— parts of the utterance (i.e., NP2
8 > NP1 and $W2 > W1$). This finding would support Levelt's (1983) account of controlled repairs by
9
10 suggesting the reallocation of resources from primary production processes to monitoring and repair
11
12 processes in later parts of the utterance. We call this the *controlled position-based account*.

13
14 3) A change to the probability of repairs covarying with the probability of errors within speakers,
15
16 irrespective of the position of the error in the sentence. This pattern would support a *controlled error-*
17
18 *based account*, and could itself show one of the following directions:

19
20 3a- A decrease in repair proportions with increasing error probability. This would point to the
21
22 sensitivity of repairs to control demands, and would specifically support a model in which primary
23
24 production and repair processes share a limited pool of resource. When primary production
25
26 conditions become more difficult, error rates increase. Since repair processes share the same
27
28 resources, without further adjustments in the system, proportion of repairs would decrease. We call
29
30 this the *fixed-resource account*.

31
32 3b- An increase in repair proportions with increasing error probability. Similar to scenario 3a, this
33
34 scenario would point to the sensitivity of the repair process to fluctuations of control. However, in this
35
36 case, when primary production processes become more difficult, the system adjustably recruits more
37
38 resources. Since these resources may not be adequate to prevent the problem to begin with, they may
39
40 manifest as increased repair rates. We call this the *adjustable-resource account*.

41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

In order to elicit errors and observe corrections, we strived to capture the natural production and monitoring processes as much as possible in an experimental environment. To this end, we chose

paradigms that require producing an utterance with a communicative intention from a conceptual (visual) event (e.g., Arnold & Nozari, 2017; Nozari & Omaki, 2018; Nozari, Arnold, & Thompson-Schill, 2014).

Participants

Twenty English-Spanish bilingual speakers (12 females; $M_{\text{age}} = 21$, $SD_{\text{age}} = 2.4$ years) were recruited from Johns Hopkins community and participated in the study in exchange for payment. The number of participants was fixed before data collection without sample size calculation, because there are no prior effect sizes available for many effects of interest in the study. All participants gave informed consent under a protocol approved by the Institutional Review Board of Johns Hopkins School of Medicine. Participants were all highly proficient in Spanish (self-reported proficiency on a 1-10 scale = 9.3, $SD = 0.9$) and English (proficiency = 9.9, $SD = 0.4$), and acquired both languages early in life ($AoA_{\text{Spanish}} = 0.0$, $SD = 0$; $AoA_{\text{English}} = 3.4$, $SD = 2.7$). Besides their high proficiency in both languages, they were dominant in English: Eighteen of them reported English as their dominant language, one of them reported Spanish, and one reported equal dominance in both. Their dominance in English was also revealed by a vocabulary test (picture naming in each of the languages), in which participants produced significantly more words in English (63.2, $SD = 1.1$ out of 65) than in Spanish (52.2, $SD = 8.2$).

Materials

The “Haunted Hotel” paradigm consisted of 224 events. Each event entailed an interaction between two colored objects (see Table 1 for the list of

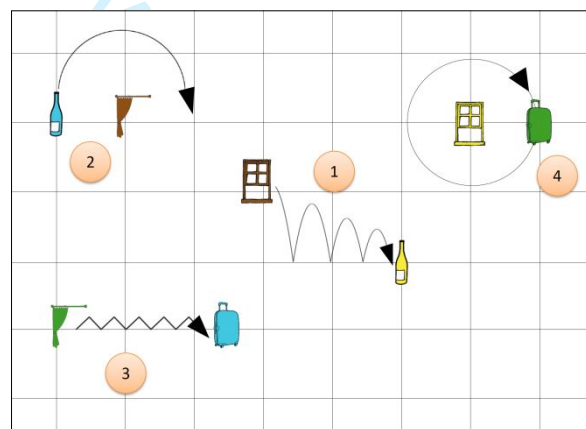


Figure 1. Example of a slide with four events that unfold in the following sequential order: 1. Bouncing towards, 2. Jumping over, 3. Zigzagging towards, and 4. Looping around. The lines show the motion path and the arrows the direction of the movement. The experiment contained 56 slides, half in English and half in Spanish blocks for a total of 224 events.

possible objects, colors, and actions, and Figure 1 for pictorial examples) in a PowerPoint slide. For example, participants saw a blue bottle disappearing behind a brown curtain, corresponding to the English sentence “The blue bottle disappears behind the brown curtain” or the Spanish sentence “La botella azul desaparece por detrás de la cortina marrón”. The events could always be described using a [NP1 + verb + NP2] structure, and NPs always consisted of a determiner, an adjective and a noun. Two sets of objects were used to create more diversity in utterances (sets 1 and 2 in Table 1). Each set (112 events) was divided into two blocks (a total of four blocks per participant), one to be described in English and one to be described in Spanish. The order of blocks and events were counterbalanced between participants, such that each participant started one set of two blocks in English and another set of two blocks in Spanish.

Spanish nouns in each set were either all feminine (set 1) or all masculine (set 2) to minimize the difficulty of choosing the correct determiner and make the English and Spanish blocks as close to each other as possible. For the same reason, to the extent possible (3 out of 4) we chose adjectives whose form did not change based on the noun’s gender in Spanish. Half of the nouns and half of the verbs were phonologically-related in English and Spanish, but this manipulation is not relevant to the current goals and will not be further discussed in this paper. Nouns, adjectives, and verbs were matched in frequency and length (in letter but not in phonemes) in English and Spanish (Appendix A).

Table 1. The linguistic materials.

Nouns	
Set 1	
English	Spanish
bottle	(la) botella
curtain	(la) cortina
window	(la) ventana
suitcase	(la) maleta
Set 2	
telephone	(el) teléfono
package	(el) paquete
mirror	(el) espejo
newspaper	(el) periódico
adjectives	
English	Spanish
green	verde
brown	marrón
yellow	amarillo
blue	azul
verbs	
English	Spanish
disappears (behind)	desaparecer (por detrás)
pass (behind)	pasar (por detrás)
produce	producir
zigzag (towards)	zigzaguar (hacia)
jump (over)	saltar (por encima)
loop (around)	rodar
bounce (towards)	brincar (hacia)
bump (into)	chocar (con)

1
2
3 Each of the four blocks was divided into 14 slides, each containing four consequent events (for a total
4 of 56 events per block, and 224 per participant). The four events on each slide each took between 2 and
5
6 of 56 events per block, and 224 per participant). The four events on each slide each took between 2 and
7
8 4 seconds and were separated from one another by a 1.5-second interval. Several aspects of the design
9
10 aimed to elicit incremental planning as the optimal strategy for completing the task: For one thing, half
11
12 of the actions were ambiguous before the completion of the event. For example, “disappearing behind”
13
14 and “passing behind” started as identical events, with the difference that in the former case the first
15
16 object never emerged from behind the second object, but in the latter case it did. This ambiguity forced
17
18 participants to wait until the action was completed to plan later parts of the sentence. Moreover, in
19
20 some cases, the second object was not clear when the action started. For example, an object may move
21
22 towards the general direction of two objects before bumping into one of them. These two
23
24 manipulations would make planning of the entire sentence at the onset of the event impossible in many
25
26 cases. At the same time, the timing of events had been carefully set using extensive pilot testing such
27
28 that speakers would run out of time if they held off on planning the first NP until the action had been
29
30 finished. Collectively, these sets of constraints strongly encouraged an incremental mode of planning: if
31
32 participants planned chunks of the sentence as actions unfolded in real time, they could speak at a
33
34 normal rate without falling behind. This mode of speaking makes substantial overlap between planning
35
36 of NP1 and NP2 unlikely.
37
38
39
40
41
42
43

44 **Procedures**

45
46 Participants completed two sessions. The first session comprised language proficiency tests in English
47
48 and Spanish. They completed the Haunted Hotel task in the second session. At the beginning of the
49
50 second session, participants completed a short orientation, introducing the objects and actions to be
51
52 seen in the experiments. They were told that all the objects were found in a haunted hotel room (the
53
54 common theme to make the semantic relations clear), that these objects would move around in the
55
56
57
58
59
60

1
2
3 room, and that their motions needed to be described to the experimenter. Participants named all
4 objects, colors, and actions in both English and Spanish and were corrected if necessary. They then
5 practiced describing each action using slides with only two objects on the screen until they could
6 produce the descriptions fluently. Finally, they completed two practice slides each with four events,
7 similar to the experimental slides. Practice was repeated if necessary, until participants could describe
8 the actions easily at the normal speech rate. Half the participants started the first set with the English
9 and the other half with the Spanish block. Since the second set consisted of new objects, the orientation
10 and practice were repeated with the new objects. All instructions were delivered in the language
11 corresponding to the first block of the set. The order of the blocks in the second set was then switched,
12 such that participants who had started the first set with an English block, started the second set with a
13 Spanish block and vice versa. Correspondingly, the language of the instructions, orientation, and
14 practice was changed in the second set, such that each participant was instructed once in English and
15 once in Spanish throughout the experiment.

16
17 During the experiment, participants produced four sentences corresponding to the four events of one
18 slide without taking a break in between (except for the 1.5 constant gap between the events). After
19 each slide, however, they could take a break, and advance to the next slide whenever they were ready.
20 This rhythm created sentence sets of four, before the flow of speech was interrupted.

21 **Statistical analyses**

22 All analyses are conducted once using non-parametric tests and once using multi-level mixed models
23 (MLMs; e.g., Jaeger, 2008) in R version 3.4.3, using the lmerTest package version 3.0-1 (Kuznetsova,
24 Brockhoff, & Christensen, 2017). Non-parametric tests have the advantage of being simple and requiring
25 few assumptions, thus removing worries about the results being unreliable due to complexities in the
26 model structure (as in MLMs) or violation of the model's basic assumptions by the data. MLMs on the
27 other hand, have the advantage of capturing a much more complex structure, e.g., items nested under

1
2
3 subjects and crossed subject-item dataframes. They can model random effects of subjects and items,
4
5 and are suitable for testing main effects and interactions in the same model. This complexity, however,
6
7 comes with issues of overfitting, lack of convergence, and sometimes violations of assumptions. We,
8
9 therefore, report the results of both methods for the analyses of interest, and look for general
10
11 convergence between the results obtained from the two methods. In fitting the MLMs, we have aimed
12
13 for the maximal random effect structure the model can handle, in keeping with the recommendations of
14
15 Barr, Levy, Scheepers, and Tily (2013). This structure was simplified for post-hoc models where the
16
17 inclusion of the full random effect caused overfitting. The random effect structure was kept consistent
18
19 between comparable models, e.g., post-hoc tests in English and Spanish. Key effects are reported in the
20
21 text. Full tables reporting the details of the MLM analyses can be found in Appendix B.
22
23
24
25

26 Results

27
28
29 For the purposes of this paper, we focus only on the NPs. Determiner errors were rare and are not
30
31 counted in this coding scheme. The rate of misses (i.e., where no description was provided for an event)
32
33 was less than 1% for English sentences and 2.5% for the Spanish sentences. This difference was not
34
35 significant ($z = 1.23, p = .22$). Any deviations from the target utterance, complete (e.g., “yellow” for
36
37 “green”) or incomplete (e.g., “yell-” for green) was counted as an error. We collected 573 errors on NPs
38
39 in total. Out of these, 54% were color and 46% noun errors. Participants made a total of 264 errors in
40
41 English ($M = 12.75; SE = 1.99$) and 309 ($M = 15.7; SE = 2.38$) in Spanish. The difference was not significant
42
43 ($z = 1.64, p = .10$). Corrections were coded as attempts to revise an uttered word, whether complete
44
45 (e.g., “yellow, no, green”), or incomplete (“yell-green”). The majority of errors (91%) were corrected
46
47 immediately, i.e., with no intervening words between the reparandum and the repair, and two thirds of
48
49 these errors were corrected before the reparandum was completed. The overall repair rate was 79% in
50
51 English and 71% in Spanish, also not significantly different from one another ($z = 0.14, p = .89$). These
52
53
54
55
56
57
58
59

1
2
3 general findings suggest that there were no obvious differences in baseline performance in English and
4
5 Spanish in our bilingual group, allowing us to compare the pattern of errors and corrections across the
6
7 two languages without worrying about large imbalances in performance potentially affecting these
8
9 patterns.
10

11 **Analysis of NP1 vs. NP2**

12
13
14 First, we examined the error rates on NP1 and NP2. The left panel in Figure 2 shows these data. Non-
15
16 parametric tests revealed significantly more errors on NP2 compared to NP1 collapsed over both
17
18 languages ($z = 3.06, p = .002$), and post-hoc tests showed the significance of the effects independently in
19
20 English ($z = 2.55, p = .011$) and in Spanish ($z = 2.15, p = .03$). The results of the MLM analyses converged:
21
22 there was a main effect of NP position ($z = 2.07, p = .038$), and no interaction between NP position and
23
24 language ($z = 0.19, p = .85$; Table B1). Post-hoc MLMs revealed significant effects of NP position in both
25
26 English ($z = 2.71, p = .007$; Table B2) and Spanish ($z = 3.026, p = .002$; Table B3). An analysis of sentence
27
28 position within a set (sentences 1, 2, 3 and 4 spoken consecutively without a break) also returned a
29
30 significant effect of NP position, consistent with the prior analysis ($z = 3.05, p = .002$), as well as an effect
31
32 of sentence position, such that error rates increased from sentence 1 to 4 ($z = 2.69, p = .007$), and a
33
34 marginal interaction between the two ($z = -1.71, p = .09$).
35
36
37
38

39
40 Next, we examined the proportion of corrected errors on NP1 and NP2. The right panel of Figure 2
41
42 shows these data. Reflecting a similar pattern to that found for errors, non-parametric tests revealed a
43
44 significantly higher proportion of corrected errors on NP2 compared to NP1 collapsed over both
45
46 languages ($z = 2.77, p = .006$), and post-hoc tests showed the significance of the effects independently in
47
48 English ($z = 2.02, p = .028$) and in Spanish ($z = 2.62, p = .009$). The MLM analysis also showed a main
49
50 effect of NP position ($z = 1.98, p = .048$), with no interaction between NP position and language ($z = -.45,$
51
52 $p = .65$; Table B4). Post-hoc MLMs revealed a significant effect of NP position in English ($z = 2.45, p =$
53
54 $.014$; Table B5) and Spanish ($z = 1.99, p = .047$; Table B6).
55
56
57
58
59
60

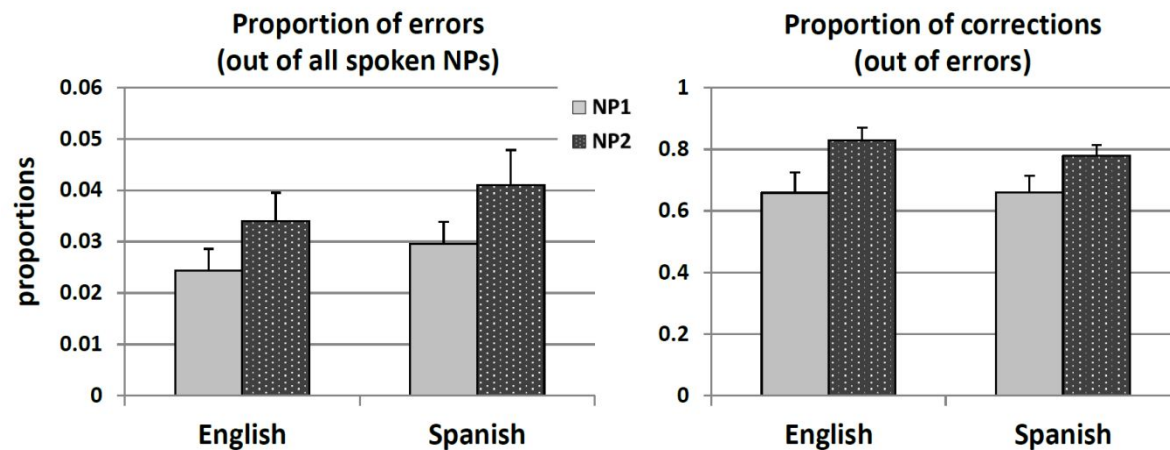


Figure 2. Proportion of errors (left) and corrections (right) on NP1 (solid light gray bars) and NP2 (polka dot black bars) in English and Spanish. Height of the bars reflects the mean of subject means and the error bars are SEs.

Discussion

The results of the error analyses provided robust evidence for the increase in the error rates from NP1 to NP2 in both languages, as expected by the semantic blocking effect. Moreover, the results showed an increase in error rates from sentence 1 to 4, consistent with a cumulative effect of semantic interference. The findings are also in agreement with Levelt's (1989) demonstration of increased error rates towards the end of the utterance, which, in both cases, can be explained by increased semantic interference. The main question of interest was whether the correction process is sensitive to the probability of committing an error. The results suggest that when error rates increase, so does the proportion of corrected errors. These findings provide an independent replication of Levelt's (1989) findings, and show a three-way relationship between position, error, and proportion of corrections. Similar to that experiment, however, it is not possible to disentangle the effect of position and proportion of errors on the proportion of corrections. The next analysis aims to do that.

Analysis of the first vs. second word within the NP

This analysis focuses on the errors and corrections on word1 and word2 collapsed over NP1 and NP2. The left panel in Figure 3 shows the error data. Non-parametric tests revealed significantly more errors on word1 compared to word2 collapsed over both languages ($z = 3.50, p < .001$). Post-hoc tests showed that the effect was significant in English ($z = 3.27, p = .001$) and marginal in Spanish ($z = 1.92, p = .055$). MLM analyses confirmed these results: there was a significant main effect of word position ($z = -3.47, p < .001$), and no main effect of language ($z = -.046, p = .96$). There was also an interaction between word

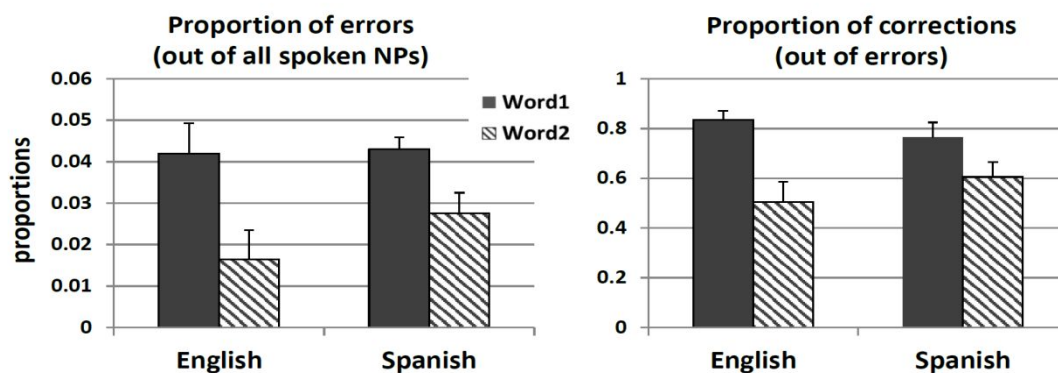


Figure 3. Proportion of errors (left) and corrections (right) on word 1 within the NP (solid black bars) and word 2 (shaded bars) in English and Spanish. Height of the bars reflects the mean of subject means and the error bars are SEs.

position and language, suggesting that the difference between error rates on word1 and word2 was more prominent in English than in Spanish ($z = 2.02, p = .044$; Table B7). Post-hoc MLM tests revealed a significant effect of word position in English ($z = -6.95, p < .001$; Table B8), and a marginal effect in Spanish ($z = -1.88, p = .06$; Table B9).

Next, we examined the proportion of corrected errors on word1 and word2 collapsed over the two NPs. The right panel of Figure 3 shows these data. In keeping with the results of the NP analysis, corrections followed a very similar pattern as errors. Non-parametric tests revealed significantly more corrections on word1 compared to word 2 ($z = 3.70, p < .001$). Post-hoc tests showed this effect to be significant in English ($z = 3.21, p = .001$) and marginal in Spanish ($z = 1.78, p = .078$). MLM analyses also showed a significant main effect of word position ($z = -4.61, p < .001$), a marginal effect of language ($z = -$

1
2
3 1.70, $p = .089$), and no interaction between word position and language ($z = 1.29$, $p = .197$; Table B10).

4
5 Post-hoc analyses using the MLM revealed significant effects in both English ($z = -4.40$, $p < .001$; Table
6
7 B11) and in Spanish ($z = -3.25$, $p = .001$; Table B12).

8 9 10 **Discussion**

11
12 As predicted, there were more errors on the first compared to the second element of the noun phrase
13
14 because of the time pressure during the NP planning. The interaction between word position and
15
16 language suggests that the difference between error rates on word1 and word2 is greater in English
17
18 than Spanish. As can be seen in Figure 3, this difference is driven by the greater reduction of errors on
19
20 word2 in English compared to Spanish. This means that participants were better able to plan the second
21
22 element of the NP while articulating the first element in the dominant (English) language. Important for
23
24 the purpose of this study, the greater likelihood of committing an error on the first compared to the
25
26 second element of the NP held in both English and Spanish, despite the fact that word1 and word2
27
28 comprised different parts of speech in these languages. The results, thus, show that later positions in
29
30 the phrase are not necessarily associated with higher error rates, i.e., that error rates do not
31
32 monotonically increase towards the end of the utterance.
33
34
35

36
37 We can now ask whether the likelihood of correction is sensitive to the position of the reparandum in
38
39 the phrase (in which case proportion of corrections should be higher on word2 compared to word1) or
40
41 to the error rates (in which case proportion of corrections should be higher on word1 compared to
42
43 word2). The results supported the second possibility. Probability of correction closely followed the
44
45 probability of errors, suggesting that it is indeed the likelihood of committing an error, and not the
46
47 position in the utterance which guides corrective behavior.
48
49
50

51 **General Discussion**

52
53
54
55
56
57
58
59
60

1
2
3 The study investigated repair to lexical errors in a task in which bilingual participants produced
4 sentences by describing simple visual events. While the goal of the current study was not to compare
5 monitoring in L1 vs. L2, it is a useful addition to the sparse empirical reports on bilingual monitoring in
6 that it reports on lexical monitoring in a homogenous, highly proficient bilingual group of speakers who
7 produce a large number of errors and repairs. A few studies that have looked at this issue are either low
8 on statistical power (e.g., Van Hest, 1996), or have focused on phonological errors and repairs (e.g.,
9 Broos, Duyck, & Hartsuiker, in press). As expected, overall performance in this group was comparable in
10 English and Spanish. One difference was a greater reduction of errors on word2 within the NP in English
11 compared to Spanish. This finding suggests better ability for parallel planning of word2 during the
12 articulation of word1 in the dominant language (Hanulová, Davidson, & Indefrey, 2011; Kroll, Bobb,
13 Misra, & Guo, 2008). The overall pattern of errors was also very similar in the two languages (Fig. 4,
14 upper panel): NP2 was significantly more error-prone than NP1, reflecting the semantic blocking effect
15 (Belke et al., 2005; Belke & Stielow, 2013; Damian & Als, 2005; Nozari et al., 2016; Schnur et al., 2006;
16 2009) at the sentence production level. Moreover, error rates on sentences 1 to 4 within a set showed a
17 monotonic increase, pointing to the cumulating nature of interference (see Belke & Stielow, 2013 for a
18 discussion). While semantic interference is often observed on response latencies in neurotypical
19 speakers, the effect manifested in error rates here because the current design requires incremental
20 planning and thus shifts the decision point on lexical retrieval towards speedy but error-prone
21 processing (Nozari & Hepner, 2018).

22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46 The pattern of within-NP errors were also similar in the two languages: there were always more errors
47 on word1 compared to word2, regardless of whether word1 was an adjective (English) or a noun
48 (Spanish; a marginal effect here). Therefore our basic assumptions about the distribution of error
49 probabilities across the sentence were verified in both languages. Importantly, the greater probability of
50 errors on word1 vs. word2 within the NP broke the monotonic relationship between error rates and
51
52
53
54
55
56
57
58
59
60

1
2
3 later positions in the sentence, allowing us to disentangle the effect of position from error probability.
4
5 Three possible outcomes were evaluated: 1) a *purely automatic account* in which repair probability is
6
7 insensitive to changes in control demands either related to position or error probability. 2) A *position-*
8
9 *based controlled account* in which repair probability increases monotonically for later positions as more
10
11 primary production processes have been completed. 3) An *error-based controlled account* in which
12
13 repair probability changes with error probability. The *inflexible-resource account* version of the error-
14
15 based controlled account predicts that repair probability would decrease with increasing error
16
17 probability. Conversely, the *adjustable-resource account* version predicts an increase in the repair
18
19 probability with increasing error rates.
20
21

22 23 **Critical findings**

24
25 We replicated the finding of Levelt (1983) by showing greater proportion of detected errors on the
26
27 later and more error-prone NP2 compared to NP1. This finding rejects the *purely automatic account*.
28
29 The results of the analysis of within-NP repairs, however, showed that it is the probability of errors, and
30
31 not the position in the sentence, that determines the probability of repair. This can be seen clearly in
32
33 Figure 4, which shows that the pattern of repairs (lower panel) is strikingly similar to the pattern of
34
35 errors (upper panel); when error rates rise, so do the proportion of corrected errors. Position, on the
36
37 other hand, is only predictive of higher repair rates when it is also associated with higher error rates, for
38
39 example, the later NP2 is also more error-prone than the earlier NP1. When later position is not
40
41 associated with higher error rates, as in the case of word2 vs. word1 within the NP, the probability of
42
43 repair tracks error rates, and not position. These findings rule out the *controlled position-based account*,
44
45 in favor of the *controlled error-based account*. More specifically, they support the *adjustable-resource*
46
47 *account*.
48
49
50
51
52
53
54
55
56
57
58
59
60

It is worth noting that even though the constraints on repairing lexical and segmental errors may be different because of the different nature of representations and the different time-constraints of lexical selection and phonological encoding processes (Nootboom, 2005); the current findings in lexical repairs have support in phonological repairs as well. Consonant errors are more likely in the onset position than in any other position in the word (e.g., Nootboom & Quené, 2015; Shattuck-Hufnagel, 1992) and onset position also happens to have higher repair rates, as well as shorter cutoff-to-repair times, than other positions (Nootboom & Quené, 2019). These findings mirror those reported in this experiment for lexical errors, and point to more efficient repair processes where the error probability is the highest in both lexical and segmental domains.

Before we discuss an account of error repair that accommodates these findings, we must address what may seem like a discrepancy with prior reports. We have previously reported evidence for a relationship between the quality of the production system and the quality of error detection/repair. In what was summarized as the conflict-based account of monitoring, we showed that damaged (Nozari, Dell, & Schwartz, 2011) or immature (Hanley et al., 2016) production systems are associated with higher levels of conflict between the target and the competitors. This high conflict gives rise to higher error rates, and also poorer distinction between error and correct trials (see Nozari & Novick, 2017, for a short review). The conflict-based model, thus, predicts a reverse relationship between the number of errors

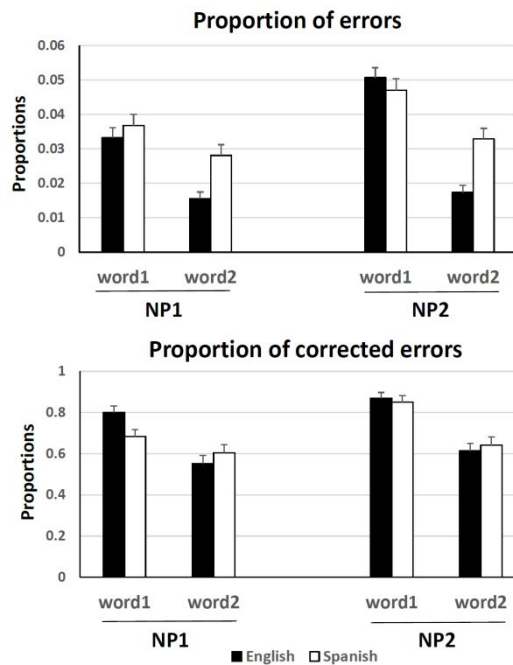


Figure 4. Summary of all the data. Proportion of errors (upper panel) and corrected errors (lower panel) for word1 and word2 in NP1 and NP2 in English (black bars) and Spanish (white bars). The figure shows the similarity between the fluctuations of error rates (decrease from word1 to word2, increase from NP1 to NP2) in the two languages, and the similarity between this pattern and the pattern of corrections.

1
2
3 (in so far as they represent the quality of the production system) and the ability to detect/repair those
4 errors, the exact opposite of what we have reported here.
5
6

7 The critical distinction here is between the general state of the production system (i.e., that which
8 cannot be much improved by implementation of control) and the temporary states (i.e., those which can
9 benefit from the implementation of control). For example, if semantic-lexical mapping has been
10 damaged (e.g., after stroke), the amount of conflict between target and non-target lexical
11 representations during each naming attempt cannot be decreased more than a certain amount no
12 matter how much control is exerted. In other words, high conflict has become a “trait” of the individual.
13
14 On the other hand, a healthy system may temporarily be in a “state” of high conflict (e.g., because of
15 time pressure during planning), which can be effectively resolved by deployment of control processes.
16
17 This trait/state difference is important in making predictions regarding the relationship between error
18 rates and proportion of repaired errors: when high conflict is a “trait”, more errors and worse
19 detection/correction performance is expected. This means that *across individuals*, one would expect a
20 negative correlation between the probability of errors and the probability of repairs. In keeping with this
21 prediction, we observed a significant negative correlation between the proportion of errors they commit
22 in Spanish and the ability to detect/correct those errors ($r = -.48, p = .034$)¹ across the individuals, adding
23 to the results of previous studies showing a similar correspondence between the quality of production
24 and the quality of monitoring/repair in individuals with aphasia and children (Hanley et al., 2016; Nozari
25 et al., 2011). *Within the same individuals*, however, fluctuations of conflict as a “state” could still lead to
26 better recruitment of control, hence the pattern of covariance between error and repair probabilities
27 reported in this study.
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52

53 **An account of repair in speech production**

54
55 ¹ This correlation is also negative in English ($r = -.27$) but does not reach significance ($p = .25$) most likely because of
56 the smaller variability in the number of errors made in English across subjects.
57

1
2
3 Generally-speaking, two classes of accounts have been proposed for repairing speech errors. The first
4
5 type views the repair process as re-initiating the production process from scratch, albeit with some
6
7 degree of priming of the repair by the reparandum (Hartsuiker & Kolk, 2001). This presumably requires
8
9 recommitting to the goal or defining a new goal to restart the production process. The second class
10
11 views the repair process as a form of revising the previous plan (e.g., Boland, Hartsuiker, Pickering, &
12
13 Postma, 2005). A version of this account was introduced in the Introduction: detection of an error
14
15 triggers a repair process which simply entails selecting the most highly activated item that was *not*
16
17 produced (see also Nootboom & Quené, 2019). Note that this process does not require resetting or
18
19 recommitting to a goal, as it picks up production from the middle of the process. This account is
20
21 reasonable, because while there is a debate on whether activation of non-target responses slows down
22
23 the production of target (e.g., Abdel Rahman & Melinger, 2009; Mahon, Costa, Peterson, Vargas, &
24
25 Caramazza, 2007; Nozari & Hepner, 2018), it is agreed by all production accounts that similar target and
26
27 non-target representations can and do become activated simultaneously. Moreover, this simple
28
29 mechanism could lead to efficient repair behavior at least in fully developed neurotypical systems,
30
31 because the target response is often highly activated even on trials where a non-target response has
32
33 been erroneously produced. Support for this claim comes from studies reporting responses that seem to
34
35 be a blend of the correct and error responses suggesting the coactivation ---and the tendency for
36
37 selection--- of both in close temporal vicinity (Goldrick & Blumstein, 2006; McMillan & Corley, 2010;
38
39 Nootboom & Quené, 2008). In weaker production systems, e.g., those of children and individuals with
40
41 aphasia, this mechanism still leads to repair attempts, although the success rate of such repairs would
42
43 be lower since competitors have comparatively high levels of activations and may be selected
44
45 mistakenly as the repair. This kind of process can account well for findings such as *conduite d'approche*
46
47 discussed in the earlier sections, and why speakers may inadvertently switch a correct response to a
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 wrong one, as repairs are done fairly automatically, as long as alternatives with high activation are
4
5 available (e.g., Nozari, 2019).
6

7 This purely automatic account, however, cannot by itself explain the relationship observed between
8
9 the error rates and proportion of corrected errors found here, and also reported in Levelt (1983) and
10
11 Nootboom and Quené (2019). The most straight forward prediction is that repair attempts in the
12
13 purely automatic account should remain fairly insensitive to error probabilities. If anything, one would
14
15 expect a decrease in the accuracy of repair outcomes, because although an automatic process may
16
17 initiate lots of repairs in error-prone situations (as in *conduite d'approche*), more competitors have
18
19 activation levels comparable to the target and can be selected as potential repairs by an automatic
20
21 system. It is thus unlikely that correct repairs are more prevalent under error-prone conditions.
22
23
24

25 An automatic mechanism can, however, be augmented by cognitive control. The increase in the rate
26
27 of repairs when speakers make more errors is compatible with greater recruitment of control resources
28
29 in these situations. More precisely, when competition is high, the monitor detects the need for greater
30
31 control, signals this need to control centers, which in turn deploy the necessary control to the part of
32
33 the system where such control is needed to resolve the conflict between competing responses. This
34
35 control is primarily put towards preventing the generation of an error (Nozari & Hepner, 2018);
36
37 however, since speakers are under time pressure, sometimes the control cannot resolve the conflict in a
38
39 timely manner, and an overt error is generated. Under such circumstances, the control system exerts its
40
41 influence to repair the errors, e.g., by suppressing the activation of the recently produced error or by
42
43 boosting the activation of the next most highly activated response so that it could overtake the error.
44
45 Since error-prone conditions are precisely those conditions that recruit the highest amount of control, a
46
47 controlled process of repair can explain the direct relationship between the probabilities of error
48
49 commission and repair.
50
51
52
53
54
55
56
57
58
59
60

1
2
3 To summarize, the empirical findings on repairing speech errors point to an underlying automatic
4 process that provides “quick and dirty” repairs. This mechanism can explain many findings such as the
5 quick timeline of repairs, as well as a rapid sequence of repair attempts in individuals with brain
6 damage, and potentially cases of unconscious repairs in production. The automatic account, however,
7 needs to be augmented with control processes to explain effects such as increased probability of error
8 correction as a function of increased error rates. Such control may be implemented strategically, i.e.,
9 speakers may become aware of the difficult parts of production and consciously divert attentional
10 control to those parts. This possibility shifts the locus of the effect from *correction* to *detection*. An
11 alternative —non-mutually-exclusive— explanation is that control is implicitly assessed and
12 implemented when necessary through incremental learning processes which help regulate production
13 (e.g., Freund & Nozari, 2018). It is a key critical question for future research to determine the nature of
14 control processes that contribute to repairs and the circumstances under which a certain type of control
15 is triggered. The answer to this question may also shed more light on why some repairs seem to be
16 readily available after error detection, while others take a while to be planned (Nooteboom & Quené,
17 2017).

18
19 Another critical question for future studies concerns the exact mechanisms by which control may
20 enhance the repair behavior in error-prone situations. One mechanism could be the triggering of an
21 activation boost (Gauvin & Hartsuiker, under review). An alternative could be stronger suppression of
22 the reparandum to give the repair a better chance of production, or a combined activation-suppression
23 account.

24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

Acknowledgment

1
2
3 We thank Michael Freund for his help with data collection. This work was supported in part by the
4
5 Therapeutic Cognitive Neuroscience Fund endowed to the Cognitive Neurology division of the Neurology
6
7 Department at Johns Hopkins University. C.D. Martin was supported by the Spanish Ministry of Economy
8
9 and Competitiveness (SEV-2015-490; PSI2017-82941-P; Europa-Excelencia ERC2018-092833) and the
10
11 Basque Government (PIBA18-29).
12
13

14 15 16 **Disclosure of interest**

17
18
19 The authors declare no conflict of interest.
20
21

22 **Data availability statement**

23
24
25 The data corpus used in this paper is currently being analyzed for (a) testing the predictions of an
26
27 implemented computational model of speech repairs, and (b) the effects directly related to bilingualism.
28
29 Once those analyses are completed and their results are submitted for publication, the data will be
30
31 made publicly available. In the meantime, they are available upon request.
32
33
34

35 **References**

- 36
37
38
39 Abdel Rahman, R., & Melinger, A. (2009). Dismissing lexical competition does not make speaking any
40
41 easier: A rejoinder to Mahon and Caramazza (2009). *Language and Cognitive Processes*, 24(5),
42
43 749–760. <http://dx.doi.org/10.1080/01690960802648491>
44
45
46 Abdel Rahman, R., & Melinger, A. (2011). The dynamic microstructure of speech production: Semantic
47
48 interference built on the fly. *Journal of Experimental Psychology: Learning, Memory, and*
49
50 *Cognition*, 37(1), 149. <https://doi.org/10.1037/a0021208>
51
52
53
54
55
56
57
58
59
60

- 1
2
3 Arnold, J. E., & Nozari, N. (2017). The effects of utterance timing and stimulation of left prefrontal cortex
4 on the production of referential expressions. *Cognition*, 160, 127–144.
5
6 <http://dx.doi.org/10.1016/j.cognition.2016.12.008>
7
8
9
10 Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory
11 hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
12
13 <http://dx.doi.org/10.1016/j.jml.2012.11.001>
14
15
16 Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a
17 semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology*, 58(4), 667–692.
18
19 <http://dx.doi.org/10.1080/02724980443000142>
20
21
22
23 Belke, E., & Stielow, A. (2013). Cumulative and non-cumulative semantic interference in object naming:
24 Evidence from blocked and continuous manipulations of semantic context. *The Quarterly*
25 *Journal of Experimental Psychology*, 66(11), 2135–2160.
26
27 <http://dx.doi.org/10.1080/17470218.2013.775318>
28
29
30
31
32 Blackmer, E. R., & Mitton, J. (1991). Theories of monitoring and the timing of repairs in spontaneous
33 speech. *Cognition*, 39, 173–194. [http://dx.doi.org/10.1016/0010-0277\(91\)90052-6](http://dx.doi.org/10.1016/0010-0277(91)90052-6)
34
35
36
37 Boland, H. T., Hartsuiker, R. J., Pickering, M. J., & Postma, A. (2005). Repairing inappropriately specified
38 utterances: Revision or restart?. *Psychonomic bulletin & review*, 12(3), 472-477.
39
40 <http://dx.doi.org/10.3758/BF03193790>
41
42
43
44 Broos, W., Duyck, W., & Hartsuiker, R. (in press). Monitoring Speech Production and Comprehension:
45 Where is the Second-Language Delay? *Quarterly Journal of Experimental Psychology*.
46
47 <https://doi.org/10.1177/1747021818807447>
48
49
50
51 Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current
52 word frequency norms and the introduction of a new and improved word frequency measure
53
54
55
56
57
58
59
60

1
2
3 for American English. *Behavior Research Methods*, 41, 977–990.

4
5 <http://dx.doi.org/10.3758/BRM.41.4.977>

6
7
8 Clark, E. V. (1978). Awareness of language: Some evidence from what children say and do. In *The child's*
9
10 *conception of language* (pp. 17-43). Heidelberg, Germany: Springer.

11
12 https://doi.org/10.1007/978-3-642-67155-5_2

13
14
15 Damian, M. F., & Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of
16
17 object names. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6),
18
19 1372. <https://doi.org/10.1037/0278-7393.31.6.1372>

20
21
22 Damian, M. F., & Bowers, J. S. (2003). Locus of semantic interference in picture-word interference tasks.
23
24 *Psychonomic Bulletin & Review*, 10(1), 111–117. <http://dx.doi.org/10.3758/BF03196474>

25
26
27 Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological*
28
29 *Review*, 93(3), 283. <http://dx.doi.org/10.1037/0033-295X.93.3.283>

30
31
32 Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., Carreiras, M. (2013). EsPal: One-stop Shopping for
33
34 Spanish Word Properties. *Behavior Research Methods*, 45, 1246-1258.
35
36 <http://dx.doi.org/10.3758/s13428-013-0326-1>

37
38
39 Endrass, T., Reuter, B., & Kathmann, N. (2007). ERP correlates of conscious error recognition: aware and
40
41 unaware errors in an antisaccade task. *European Journal of Neuroscience*, 26(6), 1714–1720.
42
43 <http://dx.doi.org/10.1111/j.1460-9568.2007.05785.x>

44
45
46 Fodor, J.A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
47
48 <http://dx.doi.org/10.7551/mitpress/4737.001.0001>

49
50
51 Freund, M., & Nozari, N. (2018). Is adaptive control in language production mediated by learning?
52
53 *Cognition*, 176, 107–130. <https://doi.org/10.1016/j.cognition.2018.03.009>

54
55
56 Gauvin, H., & Hartsuiker, R. (under revision). Towards a new model of verbal monitoring. Retrieved from
57
58 <https://osf.io/fraz9/>

- 1
2
3 Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory
4 processes: Evidence from tongue twisters. *Language and Cognitive Processes*, 21(6), 649-683.
5
6 <https://doi.org/10.1080/01690960500181332>
7
8
9
10 Hanley, J. R., Cortis, C., Budd, M.-J., & Nozari, N. (2016). Did I say dog or cat? A study of semantic error
11 detection and correction in children. *Journal of Experimental Child Psychology*, 142, 36-47.
12
13 <https://doi.org/10.1016/j.jecp.2015.09.008>
14
15
16 Hanulová, J., Davidson, D. J., & Indefrey, P. (2011). Where does the delay in L2 picture naming come
17 from? Psycholinguistic and neurocognitive evidence on second language word production.
18
19 *Language and Cognitive Processes*, 26(7), 902-934.
20
21 <http://dx.doi.org/10.1080/01690965.2010.509946>
22
23
24
25 Hartsuiker, R. J., Catchpole, C. M., de Jong, N. H., & Pickering, M. J. (2008). Concurrent processing of
26 words and their replacements during speech. *Cognition*, 108(3), 601-607.
27
28 <https://doi.org/10.1016/j.cognition.2008.04.005>
29
30
31
32 Hartsuiker, R. J., & Kolk, H. H. (2001). Error monitoring in speech production: A computational test of the
33 perceptual loop theory. *Cognitive psychology*, 42(2), 113-157.
34
35 <https://doi.org/10.1006/cogp.2000.0744>
36
37
38
39 Hartsuiker, R. J. & Moors, A. (2016). On the automaticity of language processing. In: H.J. Schmid (Ed.),
40 Entrenchment, memory and automaticity. The psychology of linguistic knowledge and language
41 learning (pp. 201-225). Berlin: De Gruyter. <http://doi.org/10.1037/15969-010>
42
43
44
45 Hartsuiker, R. J., Pickering, M. J., & De Jong, N. H. (2005). Semantic and phonological context effects in
46 speech error repair. *Journal of Experimental Psychology: Learning, Memory, and Cognition*,
47
48 31(5), 921. <http://dx.doi.org/10.1037/0278-7393.31.5.921>
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 Hester, R., Foxe, J. J., Molholm, S., Shpaner, M., & Garavan, H. (2005). Neural mechanisms involved in
4 error processing: a comparison of errors made with and without awareness. *Neuroimage*, 27(3),
5 602–608. <https://doi.org/10.1016/j.neuroimage.2005.04.035>
6
7
8
9
10 Hester, R., Simoes-Franklin, C., & Garavan, H. (2007). Post-error behavior in active cocaine users: poor
11 awareness of errors in the presence of intact performance adjustments.
12 *Neuropsychopharmacology*, 32(9), 1974. <https://doi.org/10.1038/sj.npp.1301326>
13
14
15
16 Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*,
17 13(2), 135. <https://doi.org/10.1038/nrn3158>
18
19
20
21 James, W. (1890). *The principles of psychology*. New York: Holt, Rinehart & Winston.
22
23 <http://dx.doi.org/10.1037/11059-000>
24
25
26 Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards
27 logit mixed models. *Journal of Memory and Language*, 59(4), 434–446.
28
29 <https://dx.doi.org/10.1016%2Fj.jml.2007.11.007>
30
31
32 Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children's
33 metalinguistic and repair data. [https://doi.org/10.1016/0010-0277\(86\)90040-5](https://doi.org/10.1016/0010-0277(86)90040-5)
34
35
36
37 Kohn, S. E. (1984). The nature of the phonological disorder in conduction aphasia. *Brain and Language*,
38 23(1), 97–115. [http://dx.doi.org/10.1016/0093-934X\(84\)90009-9](http://dx.doi.org/10.1016/0093-934X(84)90009-9)
39
40
41
42 Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for
43 inhibitory processes. *Acta Psychologica*, 128(3), 416–430.
44
45 <https://doi.org/10.1016/j.actpsy.2008.02.001>
46
47
48
49 Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed
50 effects models. *Journal of Statistical Software*, 82(13). <http://dx.doi.org/10.18637/jss.v082.i13>
51
52
53 Laver, J. D. (1973). The detection and correction of slips of the tongue. *Speech Errors as Linguistic*
54 *Evidence*, 132–143. <https://doi.org/10.1515/9783110888423.132>
55
56
57
58
59
60

- 1
2
3 Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.
4
5 [http://dx.doi.org/10.1016/0010-0277\(83\)90026-4](http://dx.doi.org/10.1016/0010-0277(83)90026-4)
6
7
8 Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
9
10 <http://doi.org/10.2307/1423219>
11
12 McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from
13
14 articulation. . *Cognition*, 117(3), 243-260. <http://doi.org/10.1016/j.cognition.2010.08.019>
15
16 Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by
17
18 competition: a reinterpretation of semantic interference and facilitation effects in the picture-
19
20 word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and*
21
22 *Cognition*, 33(3), 503. <https://doi.org/10.1037/0278-7393.33.3.503>
23
24
25 Moors, A., & De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychological*
26
27 *bulletin*, 132(2), 297-326. <https://doi.org/10.1037/0033-2909.132.2.297>
28
29
30 Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials
31
32 are differentially related to awareness of response errors: evidence from an antisaccade task.
33
34 *Psychophysiology*, 38(5), 752–760. <http://dx.doi.org/10.1111/1469-8986.3850752>
35
36
37 Nooteboom, S. G. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical
38
39 errors in spontaneous speech. *Errors in Linguistic Performance*. <http://doi.org/10.2307/413970>
40
41
42 Nooteboom, S. G. (2005). 10 Listening to oneself: Monitoring speech production. *Phonological Encoding*
43
44 *and Monitoring in Normal and Pathological Speech*, 167. [http://doi.org/](http://doi.org/10.4324/9780203506196)
45
46 [10.4324/9780203506196](http://doi.org/10.4324/9780203506196)
47
48 Nooteboom, S., & Quené, H. (2008). Self-monitoring and feedback: A new attempt to find the main
49
50 cause of lexical bias in phonological speech errors. *Journal of Memory and Language*, 58(3), 837-
51
52 861. <http://dx.doi.org/10.1016/j.jml.2007.05.003>
53
54
55
56
57
58
59
60

- 1
2
3 Nooteboom, S. G., & Quené, H. (2013a). Heft lemisphere: Exchanges predominate in segmental speech
4 errors. *Journal of Memory and Language*, *68*(1), 26–38.
5
6 <http://dx.doi.org/10.1016/j.jml.2012.08.004>
7
8
9
10 Nooteboom, S. G., & Quené, H. (2013b). Parallels between self-monitoring for speech errors and
11 identification of the misspoken segments. *Journal of Memory and Language*, *69*(3), 417–428.
12
13 <http://doi.org/10.1016/j.jml.2013.04.006>
14
15
16 Nooteboom, S. G., & Quené, H. (2015). Word onsets and speech errors. Explaining relative frequencies
17 of segmental substitutions. *Journal of Memory and Language*, *78*, 33–46.
18
19 <http://dx.doi.org/10.1016/j.jml.2014.10.001>
20
21
22
23 Nooteboom, S. G., & Quené, H. (2017). Self-monitoring for speech errors: Two-stage detection and
24 repair with and without auditory feedback. *Journal of Memory and Language*, *95*, 19–35.
25
26 <http://dx.doi.org/10.1016/j.jml.2017.01.007>
27
28
29
30 Nooteboom, S. G., & Quené, H. (2019). Temporal aspects of self-monitoring for speech errors. *Journal of*
31 *Memory and Language*, *105*, 43-59. <http://dx.doi.org/10.1016/j.jml.2018.11.002>
32
33
34 Notebaert, W., Houtman, F., Van Opstal, F., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error
35 slowing: an orienting account. *Cognition*, *111*(2), 275–279.
36
37 <https://doi.org/10.1016/j.cognition.2009.02.002>
38
39
40
41 Nozari, N. (2019). The dual origin of semantic errors in access deficit: activation vs. inhibition deficit.
42 *Cognitive Neuropsychology*, online March, 2019.
43
44 <https://doi.org/10.1080/02643294.2019.1587397>
45
46
47
48 Nozari, N. (2018). How special is language production? Perspectives from monitoring and control. In K.
49 Federmeier and D. Watson (Eds.), *Psychology of Learning and Motivation: Current Topics in*
50 *Language* (Vol. 68, pp. 179-23). Cambridge, MA: Academic Press.
51
52 <https://doi.org/10.1016/bs.plm.2018.08.006>
53
54
55
56
57
58
59
60

- 1
2
3 Nozari, N., Arnold, J. E., & Thompson-Schill, S. L. (2014). The effects of anodal stimulation of the left
4 prefrontal cortex on sentence production. *Brain Stimulation*, 7(6), 784–792.
5
6 <http://dx.doi.org/10.1016/j.brs.2014.07.035>
7
8
9
10 Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A
11 conflict-based account of monitoring in speech production. *Cognitive Psychology*, 63(1), 1–33.
12
13 <https://doi.org/10.1016/j.cogpsych.2011.05.001>
14
15
16 Nozari, N., Freund, M., Breining, B., Rapp, B., & Gordon, B. (2016). Cognitive control during selection and
17 repair in word production. *Language, Cognition and Neuroscience*, 31(7), 886–903.
18
19 <http://dx.doi.org/10.1080/23273798.2016.1157194>
20
21
22
23 Nozari, N., & Hepner, C. R. (2018). To select or to wait? The importance of criterion setting in debates of
24 competitive lexical selection. *Cognitive Neuropsychology*, 1–15.
25
26 <http://dx.doi.org/10.1080/02643294.2018.1476335>
27
28
29
30 Nozari, N., & Novick, J. (2017). Monitoring and control in language production. *Current Directions in*
31 *Psychological Science*, 26(5), 403–410. <http://dx.doi.org/10.1177/0963721417702419>
32
33
34 Nozari, N., & Omaki, A. (2018). Syntactic production is not independent of inhibitory control: Evidence
35 from agreement attraction errors. In *Proceedings of the 40th Annual Conference of the Cognitive*
36 *Science Society Austin, TX: Cognitive Science Society*.
37
38
39
40
41 Oomen, C. E., Postma, A., & Kolk, H. H. J. (2005). Speech monitoring in aphasia: Error detection and
42 repair behaviour in a patient with Broca's aphasia. In R. J. Hartsuiker, R. Bastiaanse, A. Postma,
43 & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech*.
44 Hove, UK: Psychology Press.
45
46
47
48
49
50 Postma, A., & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed,
51 noise-masked, and normal auditory feedback speech. *Language and Speech*, 39(4), 375–392.
52
53 <https://doi.org/10.1177/002383099603900403>
54
55
56
57
58
59
60

- 1
2
3 Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-
4 cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, *54*(2), 199–227.
5
6 <http://dx.doi.org/10.1016/j.jml.2005.10.002>
7
8
9
10 Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009).
11 Localizing interference during naming: Convergent neuroimaging and neuropsychological
12 evidence for the function of Broca's area. *Proceedings of the National Academy of Sciences*,
13
14 *106*(1), 322–327. <https://doi.org/10.1073/pnas.0805874106>
15
16
17
18
19 Schuchard, J., Middleton, E. L., & Schwartz, M. F. (2017). The timing of spontaneous detection and repair
20 of naming errors in aphasia. *Cortex*, *93*, 79–91. <https://doi.org/10.1016/j.cortex.2017.05.008>
21
22
23 Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, *42*(1–3),
24
25 213–259. [http://dx.doi.org/10.1016/0010-0277\(92\)90044-I](http://dx.doi.org/10.1016/0010-0277(92)90044-I)
26
27
28 Tydgat, I., Diependaele, K., Hartsuiker, R. J., & Pickering, M. J. (2012). How lingering representations of
29 abandoned context words affect speech production. *Acta Psychologica*, *140*(3), 218–229.
30
31 <http://dx.doi.org/10.1016/j.actpsy.2012.02.004>
32
33
34 Tydgat, I., Stevens, M., Hartsuiker, R. J., & Pickering, M. J. (2011). Deciding where to stop speaking.
35
36 *Journal of Memory and Language*, *64*(4), 359–380. <https://doi.org/10.1016/j.jml.2011.02.002>
37
38
39 Van Hest, E. (1996). *Self-repair in L1 and L2 production*. Tilburg: Tilburg University Press.
40
41 <https://doi.org/10.1075/itl.117-118.05van>
42
43
44 Van Wijk, C., & Kempen, G. (1987). A dual system for producing self-repairs in spontaneous speech:
45 Evidence from experimentally elicited corrections. *Cognitive Psychology*, *19*(4), 403–440.
46
47 [http://dx.doi.org/10.1016/0010-0285\(87\)90014-4](http://dx.doi.org/10.1016/0010-0285(87)90014-4)
48
49
50 Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing
51 structural ambiguities. *Journal of Experimental Child Psychology*, *99*(2), 75–95.
52
53 <https://doi.org/10.1016/j.jecp.2007.10.004>
54
55
56
57
58
59
60

1
2
3 Wessel, J. R., Danielmeier, C., & Ullsperger, M. (2011). Error awareness revisited: accumulation of
4
5 multimodal evidence from central and autonomic nervous systems. *Journal of Cognitive*
6
7 *Neuroscience*, 23(10), 3021–3036. <https://doi.org/10.1162/jocn.2011.21635>
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review Only

Appendix A

Lexical indices for the words used in the experiment. Lexical frequency in English was taken from the SUBTLEXus database (Brysbaert & New, 2009). Lexical frequency in Spanish was taken from the ESPAL database (Duchon et al., 2013). Stimuli are matched in frequency and letter length (all p values > 0.05), but not in the number of syllables or phonemes.

Nouns									
English	Freq/ Million	Letters	Phon	Syll	Spanish	Freq/ Million	Letters	Phon	Syll
window	86.00	6	5	2	ventana	41.57	7	7	3
suitcase	13.39	8	6	2	maleta	4.42	6	6	3
bottle	50.75	6	4	2	botella	15.62	7	6	3
curtain	10.29	7	4	2	cortina	6.87	7	7	3
mirror	24.18	6	4	2	espejo	26.27	6	6	3
newspaper	23.69	9	8	3	periódico	50.12	9	9	4
telephone	32.37	9	7	3	teléfono	33.66	8	8	4
package	22.78	7	5	2	paquete	32.90	7	6	3

Adjectives									
English	Freq/ Million	Letters	Phon	Syll	Spanish	Freq/ Million	Letters	Phon	Syll
green	72.47	5	4	1	verde	80.17	5	5	2
brown	60.12	5	4	1	marrón	6.83	6	5	2
yellow	33.80	6	4	2	amarillo/a	23.14	8	7	4
blue	102.39	4	3	1	azul	65.33	4	4	2

Nouns									
English	Freq/ Million	Letters	Phon	Syll	Spanish	Freq/ Million	Letters	Phon	Syll
to jump	69.82	4	4	1	saltar	13.10	6	6	2
to loop	6.76	4	3	1	rodear	2.16	6	6	3
to disappear	20.96	9	6	3	Desaparecer	19.86	11	11	5
to pass	108.12	4	3	1	pasar	137.95	5	5	2

to bounce	9.84	6	4	1	brincar	0.87	7	7	2
to bump	12.35	4	4	1	chocar	3.89	6	5	2
to produce	11.06	7	6	2	producir	53.75	8	8	3
to zigzag	0.45	6	6	2	zigzaguar	0.05	10	9	4

For Peer Review Only

Appendix B

Analyses using multi-level model with random subject and item effects. The dependent variable (error or correction) was coded as binary, and a logistic version of the model was used accordingly. The maximum random effect structure that the model tolerated was included, unless for post-hoc tests in which the maximum random effect included was the one tolerated by both English and Spanish post-hoc models in order to allow for direct comparison between the model results.

Table B1 - Results of the error analysis on NP1 and NP2. Data from both languages were collapsed, and an interaction term was included to investigate the possible interaction between NP position and language.

Fixed effects	Coefficient	SE	z	p-value
Intercept	-4.058	0.215	-18.853	<2e-16
NP position	0.321	0.155	2.07	.038
Language	0.387	0.217	1.789	.074
NP position x Language	0.0333	0.18	0.185	.853
Random effects	Variance			
Subject intercept	0.168			
Item intercept	0.032			
NP position subject	0.359			
NP position item	<0.001			

Table B2 - Results of the post-hoc model of error analysis on NP1 and NP2 in English.

Fixed effects	Coefficient	SE	z	p-value
Intercept	-4.124	0.235	-17.588	<.001
NP position	0.345	0.128	2.706	.007
Random effects	Variance			
Subject intercept	0.421			
Item intercept	0.238			

Table B3 - Results of the post-hoc model of error analysis on NP1 and NP2 in Spanish.

Fixed effects	Coefficient	SE	z	p-value
Intercept	-3.715	0.209	-17.74	<.001
NP position	0.355	0.117	3.026	.002
Random effects	Variance			
Subject intercept	0.444			
Item intercept	0.152			

Table B4 - Results of the correction analysis on NP1 and NP2. Data from both languages were collapsed, and an interaction term was included to investigate the possible interaction between NP position and language.

Fixed effects	Coefficient	SE	z	p-value
Intercept	0.669	0.397	1.68	.092

NP position	0.737	0.372	1.982	.048
Language	0.165	0.458	0.361	.718
NP position x Language	-0.222	0.489	-0.454	.65

Random effects	Variance
Subject intercept	0.854
Item intercept	0.456
NP position subject	0.179
NP position item	0.014

Table B5 - Results of the post-hoc model of correction analysis on NP1 and NP2 in English.

Fixed effects	Coefficient	SE	z	p-value
Intercept	0.63	0.37	1.703	.089
NP position	0.821	0.335	2.45	.014
Random effects	Variance			
Subject intercept	0.355			
Item intercept	0.569			

Table B6 - Results of the post-hoc model of correction analysis on NP1 and NP2 in Spanish.

Fixed effects	Coefficient	SE	z	p-value
Intercept	0.845	0.332	2.544	.011

NP position	0.557	0.28	1.991	.047
Random effects		Variance		
Subject intercept	0.707			
Item intercept	0.657			

Table B7 - Results of the error analysis on word position within NP collapsed over NP1 and NP2. Data from both languages were collapsed, and an interaction term was included to investigate the possible interaction between NP position and language.

Fixed effects	Coefficient	SE	z	p-value
Intercept	-3.333	0.236	-14.096	<.001
Word position	-0.893	0.239	-3.74	<.001
Language	-0.01	0.228	-0.046	.964
Word position x Language	0.551	0.273	2.018	.044
Random effects		Variance		
Subject intercept	0.444			
Item intercept	0.107			
Word position subject	0.158			
Word position item*	0.107			

Table B8 - Results of the post-hoc model of error analysis on word position in English.

Fixed effects	Coefficient	SE	z	p-value
----------------------	--------------------	-----------	----------	----------------

Intercept	-3.317	0.166	-19.961	<.001
Word position	-0.972	0.014	-6.948	<.001
Random effects		Variance		
Subject intercept	0.407			
Item intercept	<0.001			

Table B9 - Results of the post-hoc model of error analysis on word position in Spanish.

Fixed effects	Coefficient	SE	z	p-value
Intercept	-3.352	0.208	-16.15	<.001
Word position	-0.419	0.223	-1.879	.06
Random effects		Variance		
Subject intercept	0.441			
Item intercept	0.106			

Correction models on word position within NP

Table B10 - Results of the correction analysis on word position within NP. Data from both languages were collapsed, and an interaction term was included to investigate the possible interaction between NP position and language.

Fixed effects	Coefficient	SE	z	p-value
Intercept	2.058	0.331	6.225	<.001
Word position	-1.731	0.376	-4.606	<.001

Language	-0.563	0.33	-1.704	.088
Word position x Language	0.586	0.454	1.289	.197
Random effects		Variance		
Subject intercept	0.813			
Item intercept	0.024			
Word position subject	0.277			
Word position item	0.024			

Table B11 - Results of the post-hoc model of correction analysis on word position in English.

Fixedcts	Coefficient	SE	z	p-value
Intercept	1.878	0.27	6.96	<.001
Word position	-1.527	0.347	-4.399	<.001
Random effects		Variance		
Subject intercept	0.39			
Item intercept	<0.001			

Table B12 - Results of the post-hoc model of correction analysis on word position in Spanish.

Fixed effects	Coefficient	SE	z	p-value
Intercept	1.533	0.2961	5.177	<.001

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Word position	-1.11	0.3417	-3.25	.001
---------------	-------	--------	-------	------

Random effects	Variance
-----------------------	-----------------

Subject intercept	0.445
-------------------	-------

Item intercept	0.056
----------------	-------

For Peer Review Only