Published in final edited form as:

J Mem Lang. 2017 June ; 94: 75–102. doi:10.1016/j.jml.2016.11.004.

Form Overrides Meaning When Bilinguals Monitor for Errors

Iva Ivanova^{a,b,*}, Victor S. Ferreira^a, and Tamar H. Gollan^b

^a Department of Psychology, University of California, San Diego

^b Department of Psychiatry, University of California, San Diego

Abstract

Bilinguals rarely produce unintended language switches, which may in part be because switches are detected and corrected by an internal monitor. But are language switches easier or harder to detect than within-language semantic errors? To approximate internal monitoring, bilinguals listened (Experiment 1) or read aloud (Experiment 2) stories, and detected language switches (translation equivalents or semantically unrelated to expected words) and within-language errors (semantically related or unrelated to expected words). Bilinguals detected semantically related within-language errors most slowly and least accurately, language switches more quickly and accurately than within-language errors, and (in Experiment 2), translation equivalents as quickly and accurately as unrelated language switches. These results suggest that internal monitoring of form (which can detect mismatches in language membership) completes earlier than, and is independent of, monitoring of meaning. However, analysis of reading times prior to error detection revealed meaning violations to be more disruptive for processing than language violations.

Keywords

internal monitoring; language switches; within-language semantic errors; listening comprehension; reading aloud; language non-selectivity

Bilinguals are mental jugglers, and skilled ones, too: They easily switch languages when they want to but rarely switch languages by mistake (Gollan, Sandoval, Salmon, 2011; Poulis, 1999). Unintentional language switches might be rare in part because they are filtered out by an internal monitor prior to overt production (Postma, 2000). But how difficult is it to detect unwanted language switches (e.g., *perro* instead of *dog*) relative to within-language semantic errors (e.g., *cat* instead of *dog*)? Translation equivalents might be easier to detect, because they do not belong to the target language, or harder to detect, because they match the intended meaning exactly.

*Corresponding author, now at the University of Texas at El Paso Mailing address: 500 W. University Ave., Department of Psychology, University of Texas at El Paso, El Paso, TX 79902, USA, iva.m.ivanova@gmail.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

To produce an utterance, speakers need to plan their intended message at a conceptual level, access the relevant abstract lexical representations (*lemmas*) of the concepts forming the message and insert them into a structural frame, plan and retrieve the morphological, phonological and phonetic structure of the utterance and ultimately articulate it (Dell, 1986; Levelt, Roelofs, & Meyer, 1999). At some point before articulation, monitoring mechanisms ensure that these steps have been performed correctly, in accordance with the speaker's intentions and the rules of the language (Hartsuiker & Kolk, 2001; Laver, 1980; Levelt, 1989; MacKay, 1987). Evidence for the existence of an internal monitor comes from observations that some errors are corrected too quickly (within approximately 150 ms) for this to happen after listening to one's overt speech (e.g., Blackmer & Mitton, 1991). Also, errors can still be detected when overt speech cannot be monitored because of noise masking, suggesting that planned speech can be monitored internally (Lackner and Tuller, 1979; Postma and Kolk, 1992; Postma and Noordanus, 1996; for other evidence, see Oppenheim & Dell, 2008; Severens, Janssens, Kühn, Brass, & Hartsuiker, 2011). Finally, even when experimentally-elicited Spoonerisms which would result in taboo words (e.g., the Spoonerism of *hit shed*) are *not* produced, they elicit elevated galvanic skin responses, suggesting that they are generated and comprehended internally but barred from overt production by the internal monitor (Motley, Camden, & Baars, 1982).

A major proposal in models of language production is that the internal monitor operates through the comprehension system (the Perceptual Loop theory: Levelt, 1983, 1989; Hartsuiker & Kolk, 2001). Once the utterance has been assembled by the language production system, the comprehension system comprehends inner speech to perform internal inspection of the production system's output – in the same way it analyzes the utterances of others – and transfers the outcome to a central monitor. If this were so, internal monitoring should show the same perception-specific effects found during comprehension of overt speech, such as perceptual uniqueness (e.g., Marslen-Wilson, 1990). Accordingly, Özdemir, Roelofs, and Levelt (2007) showed that, similarly to auditory comprehension, *internal* phoneme monitoring was faster when the phoneme which made a word unique came early in the word than when it came late (but see Huettig & Hartsuiker, 2010; Marshall, Rappaport, & Garcia-Bunuel, 1985).

Alternatives to the Perceptual Loop theory assume instead that internal monitoring does not function through the comprehension system but is a process internal to the production system itself. Thus, a major difference between the perceptual-loop monitor and production-based monitors is that the latter have access to production-internal information before the speech plan has been fully assembled. (Pickering & Garrod, 2014; Laver, 1980; MacKay, 1987; Nozari, Dell, & Schwartz, 2011; see Postma, 2000, for a review). For example, Laver (1980) proposed multiple specialized monitors examining the output of each component of the production system (e.g., conceptual processing, lemma selection) immediately after its completion. MacKay's (1987) Node Structure Theory differs in that monitoring is not based on comparisons between intended and actual output but relies instead on statistical sensitivity to unfamiliar patterns of information flow throughout the production system. More recently, Nozari et al. (2011) proposed that speech-production monitoring is a domain-general conflict-monitoring system engaging a frontal brain region, which operates by detecting conflict between intended and produced language. A different account was put

forward by Pickering and Garrod (2014), who proposed that speakers construct forward models of their to-be-produced utterances, and monitoring involves evaluating the discrepancy between such predicted (on the basis of the forward models) utterances and the actual utterances. Internal monitoring in the present study was operationalized as perceptual loop monitoring and the study was not designed to constrain monitoring theories (but see Declerck, Lemhöfer, & Grainger, 2016). However, any monitoring theory needs to include mechanisms for detection of unwanted language switches in bilingual production, and thus our study is relevant for internal monitoring in general.

A monolingual perceptual-loop monitor would detect lexical speech errors by examining at least two aspects of the words in the pre-articulatory speech plan: their form and their meaning. Additionally, the bilingual internal monitor needs to ensure that the words in the speech plan belong to the intended language, and not to another language (which may not be understood). Broadly, the bilingual monitor could cope with this task in two ways. Unintended language switches could be detected through the examination of form, with no additional mechanism responsible for wrong-language detection. Assuming monitoring is performed over inner speech (Levelt, 1989), the monitor would detect language switches specifically by examining words' phonetic make-up. Detection would be possible because even closely related languages mismatch in certain phonetic properties (e.g., /b/ is a stop in English but a fricative or approximant in Spanish; English has vowel reduction while Spanish does not; etc). Thus, wrong-language words would at least partially mismatch in phonetic characteristics with intended-language words. Alternatively or in addition, a form-examining monitoring component could determine whether words in their entirety belong to the intended-language lexicon (e.g., the answer to "Is the word *bolsa* an English word?" is "no"). Such a mechanism – one that examines the lexical status of words – has been proposed to account for the greater likelihood of phonological substitution errors that result in existing words than in non-words (the *lexical bias effect*; Baars, Motley, & MacKay, 1975; see also Hartsuiker, Corley, & Martensen, 2005).

It is also possible that the bilingual monitor includes a separate component detecting language membership through a direct examination of words' language tags (Green, 1998; van Heuven, Dijkstra, & Grainger, 1998). Note, however, that an entirely comprehension-based monitor such as the perceptual loop monitor, which does not have access to intermediate representations in the language production system, will necessarily encounter words' form first, before being able to activate any internal properties of words such as language tags. In other words, the only way for the perceptual-loop monitor to activate the Spanish language tag of the word *bolsa* would be, at a minimum, to detect that the initial "b" is a bilabial fricative, which does not exist in English. For this reason, in the following we assume that unintended language switches are detected by the bilingual internal monitor through the examination of form (either directly or through the subsequent activation of a language tag); we return to the possibility of a form-independent language-detection monitoring component in the General Discussion.

Note that bilinguals are thought to restrict production to the target language by means of language control mechanisms, which regulate the activation of the two languages to ensure the target language is activated to a greater extent than the non-target language. According to

the most well-established proposal (Green, 1998), bilinguals inhibit the non-target language to avoid interference with the target language (but see Costa & Santesteban, 2004; Verhoef, Roelofs, & Chwilla, 2009). However, language control mechanisms may occasionally fail (as may any component of the language production system); it is then up to the monitoring mechanism to detect and correct the errors.

But how easy is it to detect unintended language switches? Translation equivalent language switches might be easier to detect than within-language semantic errors. This would be because, even though they have the intended meaning, they do not form part of the target-language lexicon, and, thus, (aspects of) their form do not exist in the intended language. On this view, the bilingual internal monitor would be more sensitive to mismatches in form than to mismatches in meaning.

Evidence for this possibility comes from a monolingual study suggesting that the internal monitor makes comparisons between intended and prepared speech at a phonological – and not semantic – level of representation. Employing a stop-signal paradigm, Slevc and Ferreira (2006) asked speakers to name pictures (when they saw the picture alone or additionally heard or saw its written name as a distractor) or withhold their response (when they heard or saw written words different from the picture name). Thus, participants did the same as they would during perceptual-loop monitoring: determine if there is a difference between a to-be-produced word and a comprehended word, and, if so, halt their production. The manipulation of interest was that the words different from the picture names (stop-signal words) could be either semantically or phonologically related to the picture names. Across five experiments, phonologically related stop-signal words made it more difficult to withhold a response, while semantically related stop-signal words had no effect (even though participants were sensitive to emotional valence). That is, it was easier for participants to detect the difference between two semantically similar but phonologically different words than between two semantically different but phonologically similar words. These results imply that internal monitoring might be primarily sensitive to form similarity with intended words: Greater form similarity makes it more difficult to detect the error, while greater form *dissimilarity* makes it easier. If so, unintended switches between languages should be easier to detect than within-language semantic errors because (aspects of) their form do not exist in the intended language.

An alternative possibility is that translation equivalent language switches, which correspond exactly to the intended meaning, would be more difficult to detect than semantic errors, which are typically less semantically related than translation equivalents. On this view, the bilingual monitor would be more sensitive to intended meaning than to intended form.

Evidence for this possibility comes from a study suggesting that the internal monitor makes comparisons between intended and prepared speech at a semantic but not a phonological level of representation. Hartsuiker, Pickering, and de Jong (2005) conducted a picture-naming study in which, on a small percentage of trials, the target picture changed to a picture with a semantically related name (e.g., strawberry – grape, Experiment 1) or a phonologically related name (e.g., lemon – leg, Experiment 2). Participants were instructed to attempt to stop naming the first picture when the display changed and name the second

picture as quickly as possible. One of the findings in this study was that participants found it more difficult to halt their response when it was similar in meaning to the target response than when it was dissimilar, but equally difficult when it was similar in form to the target response than when it was dissimilar. These results imply that the internal monitor might be primarily sensitive to meaning similarity with intended words: Greater meaning similarity makes it more difficult to detect that a word about to be uttered is an error, while greater meaning *dissimilarity* makes it easier. If so, unintended switches between languages when they are translation equivalents should be more difficult to detect than within-language semantic errors, because they correspond exactly to the intended meaning.

Another version of the account predicting that language switches would be more difficult to detect than semantic errors is that the bilingual monitor is language non-selective. In other words, it is possible that the bilingual monitor examines lexical status, and does not detect an error as long as a given form is a word in any language known by the speaker. In other words, it is possible that the bilingual monitor examines words in the speech plan by asking, “Is this a word I know?” rather than “Is this a word in the intended language?”

Assuming that the bilingual monitor uses the language comprehension system or at least that its efficiency is affected by processes which also affect comprehension, this possibility is supported by ample evidence that the bilingual comprehension system is non-selective with respect to language membership. Non-selectivity in bilingual comprehension has been demonstrated in tasks presenting words in isolation (e.g., Dijkstra, Timmermans, & Schriefers, 2000; Lagrou, Hartsuiker, & Duyck, 2011; Ng & Wicha, 2013) and in a sentence context (e.g., Dijkstra, Grainger, & Van Heuven, 1999; Lagrou, Hartsuiker, & Duyck, 2013; Marian, Spivey, & Hirsch, 2003); in reading (Duyck, 2005; Duyck, Diependaele, Drieghe, & Brysbaert, 2004; Van Hell & Dijkstra, 2002) as well as in listening (Lagrou et al., 2013); in semantically low-constraint sentence contexts (Van Assche, Duyck, Hartsuiker, & Diependaele, 2009) as well as in high-constraint ones (Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011; Titone, Libben, Mercier, Whitford, & Pivneva, 2011; but see Schwartz & Kroll, 2006); and in the non-dominant language (Jared & Kroll, 2001; Schwartz, Kroll, & Diaz, 2007; Van Assche et al., 2011) as well as the dominant one (Lagrou et al., 2013; Titone et al., 2011; Van Assche et al., 2009).

Of most relevance here, bilinguals' expectations of an upcoming word during listening comprehension are not limited to words from the target language (Lagrou et al., 2013), implying that the bilingual comprehension system might not detect a problem upon encountering non-target language words with acceptable meanings. In the study of Lagrou et al., Dutch-English bilinguals listened to sentences in both their non-dominant and dominant languages (e.g., *That man finally got a flower, and that's why he is happy*) while simultaneously looking at a four-picture display. Besides a picture of the target (e.g., a flower), the display included two types of distractor pictures: a picture whose *non-target-language* name had initial phonological overlap with the target while the target-language name was phonologically unrelated (e.g., a bottle, “fles” in Dutch); and a picture whose name was phonologically unrelated to the target in both languages (e.g., dog, “hond” in Dutch). Upon hearing the onset of the target word (the phoneme /f/), bilinguals fixated more on those distractors whose non-target-language names started with the same phoneme (fles

[bottle]) relative to the distractors whose non-target-language names started with a different phoneme (hond [dog]). Such results imply that the comprehension process in bilinguals is primarily or only concerned with arriving at the intended meaning, without much regard for the language in which this meaning is expressed. This in turn might suggest that language switches that mismatch only the intended language but match the intended meaning perfectly (i.e., translation equivalents), might be very difficult to detect by the bilingual internal monitor.

To our knowledge, only one study to date has investigated monitoring processes in bilinguals directly (see below). However, bilinguals have shown sensitivity to the lexical form (operationalized through language membership) of upcoming expected words during comprehension. Altarriba, Kroll, Sholl, and Rayner (1996) measured bilinguals' eye-movements while they read high- and low-constraint sentences in English, in which target words were sometimes substituted with their Spanish translation equivalents (e.g., *He wanted to deposit all of his money/dinero at the credit union*). High-constraint sentences produce semantic expectations about the identity of the target, and facilitation if those expectations are met. The question was whether they also produce lexical-form expectations, and interference if those expectations are violated. Crucially, translation equivalents meet the semantic expectations, but violate the lexical ones. Indeed, first fixation durations for Spanish (but not English) high-frequency targets were *longer* in the high-than in the low constraint condition; that is, in high-constraint sentences unexpected Spanish words produced interference. (This was not the case for low-frequency Spanish targets, likely because they require additional processing time sufficient to offset lexical interference.) When bilinguals read the sentences in a rapid serial visual presentation (RSVP) mode and named the targets aloud, the pattern of naming times was very similar. These results suggest that violation of form expectations disrupts processing, and that bilinguals are sensitive to the form (and language membership) of expected words during comprehension. However, it is unclear whether this form/language sensitivity is stronger or weaker relative to sensitivity to expected meaning, because Altarriba et al. did not compare translation equivalents to within-language substitutions.

This issue was investigated by Moreno, Federmeier, and Kutas (2002), who compared the processing of translation equivalent language switches to the processing of within-language synonyms. Crucially, both switches and synonyms match meaning expectations but violate form expectations for upcoming words (although note that translation equivalents are a better match of the expected meaning than within-language synonyms, which differ slightly from the expected meaning). Moreno et al. (2002) recorded event-related potentials (ERPs) while bilinguals read English sentences such as *He put a clean sheet on the ...*¹. At the end of such sentences, translation equivalent language switches (e.g. *cama* [Spanish bed] were processed in a qualitatively different way than within-language synonyms (e.g. *mattress*). Specifically, within-language synonyms elicited an enhanced negativity peaking around 400 ms post-stimulus onset (an N400), which has been associated with the processing of meaning (e.g., Kutas & Hillyard, 1980). In contrast, translation equivalent language switches

¹Moreno et al. (2002) also studied the processing of language switches and unexpected within-language synonyms in idiomatic contexts (not directly relevant here), which produced results in most ways similar to the ones for non-idiomatic contexts.

elicited a large posterior positivity (450–850 ms), which has been associated with processing unexpected or improbable events (e.g., Donchin, 1981; McCallum, Farmer, & Pocock, 1984) including improbable linguistic events (certain types of grammatical violation, e.g., Coulson, King, & Kutas, 1998). Language switches also elicited a negativity peaking around 250–450 ms post stimulus onset, but with a distribution different from N400 effects. Given the absence of an N400 effect for language switches, Moreno et al. concluded that within-language violations of lexical form (and partially of expected meaning) might be more costly to comprehend than wrong-language violations of lexical form. Applied to the current study, these results imply that translation equivalent language switches might be more difficult to detect during monitoring than within-language semantically-related words because they might disrupt processing to a lesser extent.

The only extant study directly investigating monitoring processes in bilinguals was conducted by Declerck et al. (2016). Their bilingual participants described in both French and English the transitions of a dot over a network of pictures (e.g., The dot goes left over the straight line to the duck), switching to a different-than-the-preceding language on 2/3 of the trials. In this task, bilinguals made fewer wrong-language intrusions (226; e.g., *en haut* instead of *up*) than lexical errors (382; e.g., *upper* instead of *lower*; note that these numbers were not compared directly). Additionally, bilinguals corrected wrong-language intrusions (e.g., ... *to the diable.. devil*) more often than lexical errors (*The dot goes left.. right*). In the present context, these results might suggest that language switches are easier to detect – internally, and hence avoid producing, and externally, and hence correct overtly – than within-language errors. (Note, however, that the main purpose of this study was to distinguish between the Perceptual Loop theory and the conflict-monitoring account of error detection proposed by Nozari et al. (2011). The authors assumed that language switching enhances conflict monitoring (predicting more efficient detection of intrusions on switched-language sentences than on repeated-language sentences) but hinders comprehension (predicting less efficient detection of intrusions on switched-language sentences than on repeated-language sentences). Bilinguals corrected more wrong-language intrusions on switched-language than on repeated-language sentences, incompatible with the Perceptual Loop theory according to the authors. However, it is impossible to determine the rates of intrusions corrected *prior to* overt production, and they may differ from the rates of overtly-corrected intrusions the authors report; thus, more evidence seems needed to rule out the Perceptual Loop theory.)

The present study

We investigated whether language switches are easier or more difficult to detect than within-language semantic errors during bilingual comprehension, and how this depends on the speed of processing of different types of information (form or meaning) that becomes available in the time course of error detection. For this purpose, we asked Spanish-English bilinguals to detect language switches and within-language semantic errors while listening to short stories in their dominant language English (in Experiment 1), and while reading the same stories aloud (in Experiment 2). In doing so, we assumed that internal monitoring might rely on language comprehension to detect language switches and within-language

semantic errors (Levelt, 1989), but our study has broader implications about bilingual language processing.

Error-detection tasks have been used with success to investigate a range of issues, such as reading in a second language (e.g., Hahne & Friederici, 2001; Jiang, 2007; Kaan & Swaab, 2003; Tokowicz & MacWhinney, 2005), comprehension monitoring (the real-time awareness of comprehension success; Baker, 1979, Oakhill, Hartt, & Samols, 2005; Winograd & Johnston, 1982) and text revision (e.g., Faigley & Witte, 1981; Roussey & Piolat, 2008; Sommers, 1980). The patterns of detection of lexical, syntactic and semantic errors are assumed to reflect the underlying processing associated with these different types of information. Generally, results indicate that semantic inconsistencies are more difficult to detect than form errors such as misspellings (e.g., Levy & Begin, 1984; Roussey & Piolat, 2008), presumably because they require integration of information across larger structures and induce a greater working memory load (Daneman & Stainton, 1993; Hacker, Plumb, Butterfield, Quatham, & Heineken, 1994; McCutchen, Francis, & Kerr, 1997). Accordingly, in a study of within-language bilingual error detection (focused on trained interpreters), Yudes, Macizo, Morales, and Bajo (2013) found that untrained bilinguals reading in their second language detected semantic errors such as *telephone system* instead of *immune system* less accurately (22%) than form (spelling) errors such as *kynd* or *office* (45%). What remains unclear is whether language switches would be treated by bilinguals' comprehension system in the same way as within-language form errors: Language switches have different-than-expected form, but the form is ultimately correct, except it is in the wrong language.

Bilinguals could detect unexpected language switches in several ways: through their different phonetic make-up (in listening comprehension), their absence from the target-language lexicon, or their language tags. However, all of these possibilities would require at least a partial examination of form. Since comprehension proceeds from form to meaning, words' lexical representations are activated only through phonological representations, and activation of lexical representations would be necessary to determine whether or not they belong to the target language lexicon, or whether or not they have the target-language tag.

In listening comprehension (operationalizing inner speech in Experiment 1), the first signal for a form mismatch with expected words could arrive through different-than-expected phonetic make-up and prosody. For this reason, it is important to determine whether English and Spanish, which we used in our experiments, differ sufficiently in their phonetic characteristics, phonotactics and prosody. An examination of the two systems suggests that this is indeed the case. A detailed phonetic, phonotactic and prosodic characterization of English and Spanish goes beyond the scope of this study, but here we list several of the most prominent differences between the standard variants of the two languages. English has 12 vowel phonemes which do not overlap with the five vowel phonemes in Spanish (i.e., /e/ in English is more fronted than /e/ in Spanish). English has 24 consonant phonemes, while Spanish has 19. Language-specific consonants include /ŋ/, /h/, /ð/, /z/, /ʃ/, /ʒ/, /dʒ/ for English, and /p/ and /x/ for Spanish (in English, /x/ is limited to Scottish and Irish English). In English, the phonemes /b/, /d/ and /g/ are realized as voiced stops (sounds made by the occlusion of the vocal tract which prevents airflow), while in Spanish they are most typically

realized as fricatives or approximants (sounds made by the articulators approaching each other with or without creating turbulent airflow). In English, /r/ is typically a postalveolar approximant, while in Spanish it is a flap (a sound produced by a single contact between two articulators such as the tongue and the palate) or a trill (a sound produced by vibrations between two articulators). Spanish, unlike English, does not permit consonant clusters in word-final position. English is a stress-timed language (syllables have different length), while Spanish is a syllable-timed language (all syllables have approximately the same length), and pitch variations are more pronounced in English than in Spanish. In all, there are a number of differences between English and Spanish in terms of phonetic characteristics, phonotactics and prosody. We thus assumed that, if bilingual error monitoring is sensitive to form, language switches would violate expected form to a greater extent than within-language errors.

In reading comprehension (operationalizing a loop though the comprehension system followed by articulation in Experiment 2), form mismatch could be detected through direct lexical access during reading aloud. In a dual-route model of reading (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Ziegler, 2011; Zorzi, 2010), a lexical route consists in direct access from sub-lexical orthographic information to whole-word orthographic representations, and from these, to whole-word phonology and lexico-semantic information. Accessing phonology would aid the detection of unexpected language switches in a similar way to the way they are detected in listening comprehension. Accessing lexical information would aid identifying whether or not they belong to the target-language lexicon, or to the activation of their (wrong-)language tags. In all, while processing of form might be different in reading than in listening comprehension, such processing would be necessary to identify language switches. However, since English and Spanish use the same orthographic code unlike, say, Mandarin and English (with the exception of *ll*, *ñ* and the use of diacritics in Spanish), it might be expected that bilinguals have more difficulties detecting language switches in Experiment 2 than in Experiment 1. This is because, in listening, most between-language phonetic differences would become apparent immediately (i.e., with word onset).

In our experiments, bilinguals comprehended short stories in English including sentences such as *And her shoulders always hurt because she carried too many books in her —*. We compared their speed and accuracy to detect four kinds of mismatching words at the end of such sentences: language switches which were translation equivalents of the expected completions (e.g., *bolsa* [Spanish *bag*]), language switches unrelated in meaning (e.g., *silla* [Spanish *chair*]), semantically related within-language errors (e.g., *wallet*), and semantically unrelated within-language errors (e.g., *table*). Context mismatch in meaning was created by using semantically-constraining sentences. The inclusion of the four conditions allowed us to tease apart bilinguals' sensitivity to language and meaning mismatches with expected words. If bilinguals are sensitive to language mismatches, both types of language switch (e.g., *bolsa* [bag] and *silla* [chair]) should be detected more quickly and accurately than both types of within-language error (e.g., *wallet* and *table*). If bilinguals are sensitive to meaning mismatches, words completely unrelated to the expected completions (e.g., *silla* [chair] and *table*) should be detected more quickly and accurately than semantically related words (e.g., *bolsa* [bag] and *wallet*), regardless of language membership.

The four conditions also allowed us to observe the relative time-course of monitoring processes. First, monitoring for form acceptability (which we assume would detect language switches) may be independent of, and complete more quickly than, monitoring for meaning. If so, both types of language switch (*bolsa* [bag] and *silla* [chair]) would be detected first and equally quickly and accurately, because both deviate in form from expected completions and meaning monitoring will not have yet completed. When monitoring for meaning is subsequently completed, within-language errors would also be detected, but later and less accurately than language switches. Also, related within-language errors (*wallet*) should be detected more slowly and less accurately than unrelated errors (*table*) because the greater meaning overlap with expected completions would make them more difficult to detect. The predictions for this possibility are plotted in Figure 1, Panel A.

An alternative possibility is that monitoring for meaning completes before monitoring for form (because arriving at the expected meaning is the ultimate goal of comprehension). If so, both unrelated words (*silla* [chair] and *table*) should be detected first and equally quickly and accurately, because both mismatch in meaning to the same extent, and form monitoring would not yet have completed. Related within-language semantic errors (*wallet*) would be detected more slowly and less accurately than unrelated words because of the partial meaning overlap. Subsequently, form monitoring would complete, and translation equivalent language switches (*bolsa* [bag]) would be detected, but most slowly and least accurately. Predictions for this possibility are plotted in Figure 1, Panel B. Note that both possibilities explained so far predict a statistical interaction between the language and relatedness factors.

A third possibility is that detection of form mismatches interfaces, and operates in parallel, with detection of meaning mismatches. If so, the words mismatching in *both* form and meaning (*silla* [chair]) should be detected first and most accurately, followed by words mismatching only in form (*bolsa* [bag]) or only in meaning (*table*), and words both matching expected form and with some meaning overlap (*wallet*) should be slowest and least accurate. In this possibility, there should be no statistical interaction between language and relatedness. Language switches are overall faster because, collectively, they mismatch in more parameters (both mismatch in form and one also mismatches in meaning) than within-language errors (only one clearly mismatches in meaning). Predictions for this possibility are plotted in Figure 1, Panel C.

In Experiment 1, bilinguals detected switches and errors while listening to recorded stories (which, according to the Perceptual Loop Theory, should engage processes in some ways similar to internal monitoring). But monitoring for errors while listening to someone else is also different from internal monitoring because, during language production, a speaker knows her intended utterance in advance, and because she must ultimately engage in articulation. Experiment 2 was thus designed to better approximate internal monitoring: Bilinguals performed the same switch and error detection task with the same stories as in Experiment 1, except they read the stories aloud. Reading aloud allows for some advance processing because upcoming words are typically recognized approximately 500 ms before they are articulated (the *eye-voice span*: Buswell, 1922; Inhoff, Solomon, Radach, & Seymour, 2011). The eye-voice span is sensitive to semantic and syntactic information (Vázquez, Glucksberg, & Danks, 1977) as well as homographs (Buswell, 1922), indicating

that it allows processes of semantic integration (for example, the reader needs to decide which of the two meanings of a homograph fits the context). Thus, reading aloud (unlike listening comprehension) might better resemble internal monitoring during speech production because bilinguals would have to actually speak, and because they would know ahead of time aspects of the language they are going to produce. A potential drawback is that the perceptual-loop monitor is assumed to operate over inner speech (phonetic code) rather than over orthographic code.

This choice of task in Experiment 2 provided a further measure of interest. In addition to measuring speed and accuracy of switch and error detection via button-press responses (as in Experiment 1), we additionally sought to reveal the types of information governing advance processing during the eye-voice span. Note, however, that we assumed that such advance processing would be governed by different mechanisms than error detection. To this aim, in Experiment 2 we measured the time bilinguals took to read aloud the five words preceding the mismatching words, henceforth *articulation times* (the eye-voice span is at most four to six words: Inhoff et al., 2011).

Experiment 1: Monitoring for Switches and Errors During Listening

In this experiment, Spanish-English bilinguals listened to short stories in English. They were asked to pay attention to the content, and to press a button whenever they detected Spanish words (language switches) or English words that did not fit the context (within-language errors). We measured speed and accuracy to detect the switches and errors.

Method

Participants—Forty-eight Spanish-English bilinguals who were undergraduates at the University of California, San Diego (UCSD), participated for course credit. Eight participants were excluded because they were Spanish-dominant; the data of 40 English-dominant participants were retained for analyses. The bilinguals in our study were thus all switched-dominance bilinguals – they learned Spanish first, but subsequently became more dominant in English. Participant characteristics are summarized in Table 1. Language dominance was determined using self-rated language proficiency (participants were classified as Spanish-dominant if they rated their proficiency in Spanish as higher than their proficiency in English).

Materials—We created sixteen single-paragraph short stories in English ($M = 296$ words/21 sentences, $SD = 18$ words/0 sentences; the stories are listed in Appendix A). Each story included eight critical sentences designed to be semantically-constraining (e.g., And her shoulders always hurt because she always carried too many books in her bag). The intended completions across the 128 critical sentences were 64 high-frequency English words ($M = 125.92$, $SD = 119.94$, frequency per million from N-watch, Davis, 2005); two critical sentences in two different stories had the same intended completions. The intended completions for the eight critical sentences in each story were semantically unrelated, to avoid semantic priming effects. Cloze probabilities for the intended completions ($M = .72$, $SD = .28$) were collected from 48 native English speakers who did not participate in the main

experiment. Each of these participants saw eight stories with blanks in place of the critical words, and filled in the blanks with the first word that came to mind.

When a word was intended in a given sentence (e.g., *bag* above), it was never presented. Instead, each critical sentence ended with a word from one of four conditions: (1) Related language switch (Spanish translation equivalent), (2) Unrelated language switch (Spanish semantically unrelated word), (3) Related within-language error (English semantically related word), and (4) Unrelated within-language error (English semantically unrelated word). The four conditions were created as follows. Condition (1) was created by translating the intended completions of critical sentences (e.g., *bag*) into Spanish (e.g., *bolsa*). Note that the 64 English words serving as intended completions were chosen such that they formed semantically related pairs (e.g., *bag* – *wallet*); the other word in a pair formed condition (3). The semantic similarity between the words in a pair was determined by whether condition (3) was a noticeable error in its respective critical sentence (if the word intended as a within-language semantic error was a very near synonym of the intended completion, it would have been virtually impossible to notice that this word was an error).

To create conditions 2 and 4, the two semantically related pairs were combined into a quadruplet, such that there was no semantic relation across pairs; e.g., *bag-wallet* + *chair-table*). Condition 2 was created by translating a word from the second pair (e.g., *chair*) into Spanish (*silla*); condition 4 was created with the remaining word from the quadruplet (e.g., *table*). All four words in a quadruplet served as intended completions in different sentences (though, as noted, they were never presented to subjects as such). The degree of phonological overlap of intended completions with errors and switches was minimal and was similar across conditions (condition 1: 4/64 words; condition 2: 4/64 words; condition 3: 4/64 words; condition 4: 6/64 words [all p s > .9]); in all cases, it was limited to the first phoneme. Each quadruplet (together with the English words' Spanish translations) formed one item (see Table 2 for an example). There were 16 items in total (a list is provided in Appendix B). All Spanish words within an item had the same gender (e.g., *bolsa* [bag], *cartera* [wallet], *silla* [chair], and *mesa* [table] are all feminine) and were non-cognates. Each condition occurred between one and three times within a single story; across all 16 stories, each condition occurred an equal number of times (32).

Due to the multiple constraints for item selection, some items contained both count (e.g., *table*) and mass nouns (e.g., *love*), which rendered certain words ungrammatical in the sentences in which they appeared – in addition to semantically anomalous and/ or belonging to the wrong language (e.g., *...so every day at breakfast, they drank a lot of window*). Analyses excluding the five such items produced an identical pattern of results with a few exceptions not crucial for our main conclusions. These exceptions are reported in Appendix D.

Four versions of each story were created, such that a sentence designed for a given intended word (e.g., *bag*) ended instead in words belonging to each of the four conditions ((1) *bolsa*; (2) *silla*; (3) *wallet*; (4) *table*). Participants were randomly assigned across versions. Each participant was presented with all 16 stories and saw each of the 64 English words and each of their translations only once, in different conditions (see Table 3); thus, each participant

saw all 128 words. The stories were constructed so that each participant always saw both a word from the related conditions (either (1) or (3)), and its unrelated control (respectively (2) or (4)). For this purpose, Stories 1-8 were paired with Stories 9-16 (e.g., Story 1 with Story 9, Story 2 with Story 10, etc.) so that the stories in each pair contained sentences designed to end with the same eight words, in the same order (Tables 2 and 3)². Items were distributed randomly across the first 8 paragraphs (e.g., Item 4 occurred in Stories 1, 2, 4, 5, and their respective pairs, Stories 9, 10, 12, 13). Between one and two filler sentences occurred at the beginning and end of each story and between the sentences containing the critical words. Critical words were thus separated by at least 21 words (Mean = 33 words; Max = 54 words). At least some fillers were of high-constraint (e.g., Veronica and Richard were safe, but everything else burned down; see Appendix A for all stories); thus, not all high-constraint sentences in the stories resulted in expectation violations.

A Spanish-English bilingual research assistant recorded the four versions of each paragraph, as well as a fifth version containing only intended completions. The words from the four conditions were then spliced into the version containing the intended completions using Audacity (the version with intended completions was never presented to participants). The splices contained between one and three words (most frequently determiners and prepositions) preceding the critical words, to avoid attracting attention to the critical words in case the splice was noticeable, as well as to avoid splices of continuous speech and prevent critical words sounding strange because of coarticulation effects.

Procedure—At the beginning of the experiment, participants were given the following instructions:

In this experiment, you will listen to short stories. Your job will be to answer questions about these stories, so please listen carefully. The stories are mostly in English, but there are also some Spanish words here and there. Also, the stories mostly make sense, but there also are some errors here and there. Whenever you encounter a Spanish word, or hear a word in English that doesn't quite make sense, press the red button to indicate that you noticed the language change or the error.³

An additional (shorter) story served as practice. Three free-response comprehension questions asking for specific content (e.g., *What did the man offer Alexandra?*) were shown on the computer screen after the end of each story, and participants' spoken responses were recorded and subsequently coded for accuracy (the comprehension questions for all stories are reported in Appendix C). The stories were presented with Psyscope (Cohen, MacWhinney, Flatt, & Provost, 1993), in a different random order for each participant. After the end of the experiment proper, participants provided English translations of the 64

²The sentences in Stories 1-8 were higher in cloze probability ($M = .77$, $SD = .27$) than the sentences in Stories 9-16 ($M = .67$, $SD = .28$) [$t(63) = 2.80$, $p = .007$]. Additional analyses only on the data from Stories 1-8 (fully counterbalanced) produced an identical pattern of results with three exceptions, reported in Appendix D.

³This order of mention of the two languages, which was the same for all participants in both Experiments 1 and 2, raises the possibility that participants prioritized the search for language switches over the search for within-language errors. This possibility is supported by the results of Experiment 1 (language switches were detected more quickly and accurately than within-language errors), but is undermined by the results of Experiment 2 (unrelated within-language errors were detected as quickly as language switches, suggesting that detection of the latter was not prioritized).

Spanish words (to ensure that they knew their meanings, since the dominant language for most participants was English), and completed a language history questionnaire.

Data analysis—We analyzed response times (RTs) for detection of language switches and within-language errors, and the failures to detect those (henceforth *button-press misses*), respectively with linear and logistic mixed-effects regression models (LMER; Baayen, 2008) in R. The models had language (language switch, within-language error) and relatedness (related, unrelated) as fixed predictors. We ran two additional models per dependent variable in each experiment to target the simple effects of relatedness and of language (by removing the respective main effect from the model). All models had the maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013), unless otherwise specified. Outliers (2.5 standard deviations above or below the mean) were excluded from analyses. Extra button presses (those preceding critical words, falling in between two critical words, or following another one in close succession) were excluded from analyses as false alarms.

Results

4.99% of all button presses were excluded as false alarms, and 3.42% of the remainder as outliers (2.5 standard deviations above or below the mean). Statistical analyses are summarized in Table 4, and RTs are plotted in Figures 2a (RTs) and 2b (button-press misses). Bilinguals were both slower and less accurate to detect within-language errors than language switches (language was a significant predictor), indicating that they were sensitive to the target language and thus likely to the form (broadly construed) of target words. Bilinguals were also slower and less accurate to detect related than unrelated words (relatedness was also a significant predictor), indicating that they were also sensitive to the intended meaning.

As explained in the Introduction, of particular interest was also whether the two predictors interacted. In the RT data, the interaction between language and relatedness did not reach significance; the models targeting simple effects showed an effect of relatedness for within-language errors but not language switches (see Table 4).

For button-press misses, the interaction was again not significant, and simple effects showed an effect of relatedness for within-language errors as well as language switches (though note that the former [$z = 6.70$] was numerically (but not statistically) larger than the latter [$z = 1.97$]; see Table 4). Taken together, these results hint at an interaction, but are inconclusive.

Mean comprehension accuracy was 67% (SD = 11%), possibly because of memory limitations given the difficulty of the materials (questions were free-response and were presented only at the end of each story; in all stories, new ideas were introduced every 2-3 sentences, to avoid semantic priming effects between expected completions of critical sentences).

Analyses of the post-experiment vocabulary check indicated that participants produced the intended English translations of the Spanish words included in the experiment on 91% of the cases (96% if near synonyms such as *cup* for *glass* were counted as correct). Analyses excluding the three words producing less than 80% correct translations which were not near

synonyms of the intended completions (*muro* [wall], *cuadro* [painting] and *cerca* [fence]) produced an identical pattern of results.

Discussion

In this experiment, bilinguals detected language switches more quickly and accurately than within-language errors, indicating their sensitivity to the language membership – hence form – of expected words. Additionally, bilinguals were slower and less accurate to detect semantically-related within-language errors than unrelated within-language errors, indicating that they were also sensitive to the meaning of expected words. In fact, related within-language errors were not only close in meaning to the expected meaning, but they also generally matched in expected form (to the extent that they belong to the target-language lexicon, and have the phonetic make-up of items from the target language). Thus, in accordance with our predictions, they were the most difficult to detect by both meaning- and form-monitoring processes.

Less clear was whether the monitoring processes analyzing form and meaning were independent of each other. We predicted that, if meaning monitoring completes after form monitoring, we should see an interaction between language and relatedness (as plotted in Figure 1, Panel A). However, this interaction was not significant for either RTs or button-press misses (although it trended in the predicted direction). Simple effects were also inconclusive: Bilinguals detected related and unrelated language switches equally quickly, suggesting that form monitoring operates first and is insensitive to meaning. However, bilinguals missed related language switches significantly more often than unrelated ones, suggesting that form monitoring is not completely insensitive to meaning.

One potential criticism of Experiment 1 is that listening comprehension does not sufficiently approximate internal monitoring processes. Specifically, it could alter the extent to which form and meaning monitoring processes overlap in time. During listening comprehension, one has to analyze form and meaning for mismatches with expected words while *simultaneously* interpreting the text to determine what the expected words are. Since deriving expectations for upcoming words takes time and the meaning-monitoring process can begin only after this process has completed, the meaning monitor would necessarily be slower than the form monitor. Thus, in listening comprehension, both types of within-language error (only mismatching in meaning but not form) would be detected more slowly than language switches. However, it is unclear whether this would hold for internal monitoring processes because they should already have access to intended conceptual content and therefore might not need to run in parallel with interpretation processes. If so, during internal monitoring, form and meaning could be examined in parallel.

Furthermore, listening comprehension does not engage speech production, which in normal circumstances would follow internal monitoring (and ultimately is the reason why internal monitoring is performed). Preparing a phonetic code and translating it into motor movements might interact with error monitoring to alter the time-course of different monitoring processes. Evidence for articulation-related processing was provided by Kolers (1966): Bilinguals read mixed-language and single-language paragraphs equally quickly in

silent reading, but took *longer* to read the mixed-language than the single-language paragraphs when reading aloud (but see Altarriba et al., 1996).

In Experiment 2, we sought to better approximate the relative timing of monitoring processes relative to the concurrent planning of overt speech. For this purpose, we asked bilinguals to monitor for switches and errors during reading aloud, which simultaneously engages the speech production system (Gollan, Schotter, Gomez, Murillo, & Rayner, 2014), and provides speakers with information about upcoming words before articulation (Buswell, 1922).

Experiment 2: Monitoring for Switches and Errors During Reading Aloud

In this experiment, Spanish-English bilinguals read aloud the same short stories as in Experiment 1, and were asked to pay attention to the content and to press a button whenever they detected language switches or within-language errors.

We anticipated the following possible differences in error detection during reading aloud relative to error detection during listening (Experiment 1). Preparation for articulation during reading aloud might interact with error monitoring to produce a qualitatively different pattern from the one we observed in Experiment 1. Alternatively, reading aloud might influence the pattern of results only quantitatively. Specifically, with advance processing time, form monitoring might complete even earlier than meaning monitoring. If so, related and unrelated language switches would be detected not only equally quickly (as in Experiment 1) but also equally accurately (unlike Experiment 1). Also, with advance processing time, information about expected meaning might be available earlier and drastic meaning mismatches would be more obvious earlier on. If so, unrelated within-language errors would be detected more quickly relative to language switches than in Experiment 1. Altogether then, Experiment 2 might reveal a more robust interaction between language and relatedness than Experiment 1, with no relatedness effect for language switches, and a smaller, or absent, language effect for unrelated words. However, we would not expect notable differences for the semantically-related within-language errors relative to the other data points: They should still “fool” both the meaning and form monitoring processes and remain most difficult to detect. It is furthermore possible that language switches stand out less in this experiment than in Experiment 1, because phonetic differences between English and Spanish might be more prominent than orthographic differences (the two languages use largely the same alphabet). If so, language switches should be harder to detect in this experiment relative to Experiment 1. Lastly, if upcoming words are processed ahead of time (because of the eye-voice span), we should observe overall faster RTs than in Experiment 1.

This experiment further allowed us to obtain information about how the errors and switches affected natural language comprehension. We assume this is a different process than error detection insofar as its ultimate goal is the interpretation of meaning as opposed to the examination of specific features of actual against expected input. For this purpose, in this experiment we also measured the time bilinguals took to articulate the five words preceding (but not including) the switches and errors. These articulation times should reflect the processing of switches and errors related to integration with preceding context (Buswell,

1992; Vázquez et al., 1977) and preparation for articulation (Kolars, 1966). We assumed that slower articulation times would index processing difficulties. Note, however, that greater difficulties would ensue in this early processing window if it is relatively easy to perceive a mismatch between expected and encountered words (harder-to-notice difficulties would presumably be evident outside of the five-word window reflected in this measure). Thus, articulation times would be informative about which error or switch type would pose a particular challenge for processing before the overt detection response (button-press). If these were words mismatching in expected form, articulation of the five preceding words should be slower before language switches than before within-language errors. If these were words mismatching in expected meaning, articulation of the five preceding words should be slower before unrelated than before related words. In addition, articulation times allowed us to observe if the automatic processes governing comprehending and producing language match the controlled processes of overt error detection. If so, articulation times should pattern similarly to button-press response times; if not, we should observe notable differences in the patterns for the two measures.

We measured articulation times over a five-word window, which is much longer (1498ms on average) than the average eye-voice span (~500ms). We chose a longer time window to control for individual differences in eye-voice span and thus be sure that all bilinguals began to process the switches and errors within this window. The five words themselves were identical across conditions, hence processing these words alone should not produce articulation-time differences between conditions; such differences should arise only from any differences in the pre-articulation processing of the switches and errors. To address between-subjects' baseline differences in articulation times (given that each participant saw a particular five-word chunk only once), our LMER models had both subjects' and items' random intercepts and slopes (see the Data Analysis section of Experiment 1, which also applies to the data analyses of Experiment 2).

Method

Participants—Forty-eight switched-dominance Spanish-English bilinguals at UCSD participated for course credit. Of these, ten bilinguals were excluded (eight were Spanish-dominant, one misunderstood the instructions, and one had not learned Spanish until age 16); the data of 38 participants was retained for analyses (see Table 1 for participant characteristics). Participants' characteristics did not differ between Experiments 1 and 2 (all $ps > .15$).

Materials and procedure—These were the same as in Experiment 1, with the following exceptions. The sixteen stories were presented in written form on the computer screen, one at a time, in Courier 24-point font, across 17 lines of text each consisting of an average of 17 words. Errors and switches never occurred at the end of a line on the screen. Participants were asked to read the stories aloud (including errors and switches, which they were supposed to read normally), and concurrently press the red button on the button box as quickly as they could when they saw a Spanish word or a word in English that did not make sense. This experiment was administered together with another involving picture naming in

English and Stroop color-naming in Spanish. Order of experiments was counterbalanced and did not influence the data patterns we report (see Appendix D).

Coding—Bilinguals were recorded while reading the paragraphs aloud. Speech was recorded in the left channel, and button presses were marked by beeps in the right channel. Timestamps for the button presses, beginnings of error and switch words and beginnings of the fifth word preceding the errors or switches were extracted with Audacity by three trained research assistants. One research assistant coded stories 3,4,11 and 12 for participants 1-14, and two further research assistants coded the remainder, counterbalanced between them for participants and stories; thus each participant's data was coded by either two or three research assistants. The two-way interclass correlation coefficient was .98 between all three research assistants (for participants 1-14), and .95 between the two who coded the bulk of the data (for all participants).

Results

In RT analyses, 3.61% of all button presses were excluded as false alarms; of the remainder, 4.12% were excluded as outliers (2.5 standard deviations above or below the mean).⁵ In articulation times analyses, 5.26% were excluded because of disfluencies or reading errors on the critical words or five preceding words; of the remainder, 2.80% were excluded as outliers. Statistical analyses for the predictors of interest are summarized in Table 5, and RTs and button-press misses are plotted in Figures 3a and 3b. As in Experiment 1, bilinguals were significantly slower and less accurate to detect within-language errors than language switches (the language predictor was significant), and slower to detect related than unrelated words (the relatedness predictor was also significant). Of particular interest, the interaction between language and relatedness was significant for both RTs and misses; simple effects indicated an effect of relatedness for within-language errors but not for language switches. Simple effects also indicated that unrelated within-language errors were detected as quickly as unrelated language switches, but were missed more often.

We furthermore compared the rate of button-press misses for language switches between Experiment 1 and Experiment 2, to determine if language switches were harder to detect during listening or during speaking (RTs cannot be compared directly because, in Experiment 2, they are measured from the onset of articulation and not from the onset of processing). For this purpose, we ran a model on the misses of language switches only, with relatedness and experiment as fixed predictors, random intercepts, and random slopes for relatedness, for both subjects and items. This model indicated that there was a higher rate of misses in Experiment 2 than in Experiment 1 [$Estimate = -1.47$, $SE = .25$, $z = -5.92$, $p < .001$].

Articulation times of the five words preceding the errors and switches are plotted in Figure 3c. This measure exhibited a pattern that was considerably different from the results reported thus far. Bilinguals *slowed* their reading aloud to a greater extent before language switches than before within-language errors (the language predictor was significant). They also

⁵Since pressing buttons while reading aloud might have made the task more challenging, making outlier exclusion unwarranted, we reanalyzed the data without removing outliers; all effects and interactions reported below remained unchanged.

slowed their reading aloud to a greater extent before unrelated than related words (the relatedness predictor was significant). The two predictors also significantly interacted, but in a different way than for the RTs and misses. Specifically, articulation times were slowest before unrelated language switches, faster before unrelated within-language errors (and the two differed from one another) and fastest before related words from both languages (and the two did not differ from one another).

Mean comprehension accuracy was 73% (SD = 10%). Analyses of the post-experiment vocabulary check indicated that participants produced the intended English translations of the Spanish words included in the experiment on 90% of the cases (94% if near synonyms such as *cup* for *glass* were counted as correct). Analyses excluding the four words eliciting less than 80% correct translations which were not near synonyms of the intended completions (*cadena* [chain], *cerca* [fence], *cuadro* [painting], *muro* [wall]) produced an identical pattern of results, except, in the articulation times analyses, language became marginally significant [$p = .08$].

Discussion

In this experiment employing reading aloud, bilinguals detected language switches more quickly and accurately than within-language errors, thus showing sensitivity to language – and form – mismatches with expected sentence completions. Bilinguals also detected unrelated words more quickly and accurately than related words, thus showing sensitivity to meaning mismatches with expected completions. Of greatest interest, in this experiment there was clear evidence for a time-course separation between the completion of form and meaning monitoring, in that the interaction between language and relatedness was significant for both RTs and misses. Specifically, related and unrelated language switches were detected first and equally quickly, unrelated within-language errors were detected as quickly as, and slightly less accurately than, language switches, and related within-language errors were detected considerably more slowly and less accurately than all other error types. This pattern suggests that form and meaning monitoring run along different time-courses, with form monitoring completing first without sensitivity to meaning, and meaning monitoring completing only after form monitoring has terminated (Figure 1, Panel A). Much faster RTs than in Experiment 1 confirmed that our advance processing manipulation had an effect (in some cases, button presses occurred even before bilinguals overtly produced the language switches or error words).

In this experiment, bilinguals missed language switches more often than in Experiment 1 employing listening comprehension, possibly because Spanish differs from English phonetically more than orthographically. This between-experiment sensitivity to form similarity of language switches with expected words thus supports the assumption that language switches in both experiments are detected by a form-monitoring component.

Articulation times for the five words prior to overt production of the errors or switches, which we used as an index of advance processing during natural language comprehension in reading aloud, presented a very different picture than the button presses and misses. Assuming that slower articulation times reflect processing difficulties which become apparent early on, we observed that neither related within-language errors nor related

language switches (translation equivalents) seemed to cause such difficulties. This result reveals a sensitivity to meaning early on in processing, in that completely unrelated words elicited greater slowing, without much consideration of form or language membership. Interestingly, this initial sensitivity did not seem to qualitatively influence the pattern of overt error detection, which was similar to the one during listening comprehension without advance processing (observed in Experiment 1).

It was unclear why unrelated language switches slowed reading aloud more than unrelated within-language errors. This effect could be due to longer preparation for articulation for words with non-target-language phonetic make-up (unrelated switches) than for words with target-language phonetic make-up (unrelated within-language errors; Kolers, 1966). Note that, if this were so, related language switches (which also have non-target-language phonetic make-up) might have caused such slowing as well. But if non-target-language phonology slowed reading down before related switches (but not before related within-language errors), the *meaning* of these switches would in fact have been processed faster than the meaning of the errors. Consistent with this possibility is the fact that related switches overlap in meaning with expected words to a greater extent than related within-language errors, which have some, but not complete, semantic overlap.

General Discussion

In this study, we asked if language switches or within-language errors are harder to detect during internal monitoring in bilinguals, and what types of information determine the speed and accuracy of switch and error detection in the time-course of processing. We operationalized internal monitoring through overt monitoring during comprehension. Bilinguals thus listened to stories in English (in Experiment 1), or read the same stories aloud (in Experiment 2), and were asked to pay attention to the content and to detect language switches and within-language errors. Both the switches and errors could be semantically-related (related language switches were identical in meaning) or unrelated to the expected words they were substituting. In both experiments, bilinguals detected language switches more quickly and accurately than within-language errors. This result evidences a form-monitoring process because language switches differ in form from expected completions more than within-language errors do. Also in both experiments, bilinguals were slowest and least likely to detect related within-language errors. This result evidences a meaning-monitoring process because greater meaning similarity (when the errors' form matched expectations for target-language form) resulted in greater difficulty of detection.

Furthermore, when the overt monitoring situation presumably better approximated internal monitoring (Experiment 2), both types of language switch were detected equally quickly and accurately whether or not they were related in meaning to the expected target, whereas related within-language errors were detected more slowly and less accurately than unrelated within-language errors. In Experiment 1, this pattern was statistically less conclusive but also apparent. In the context of our experiments, this pattern suggests that the form- and meaning-monitoring processes were independent of each other (predictions plotted in Figure 1, Panel A). This is because, if the form- and meaning monitoring processes interacted, related language switches should have been detected more slowly and less accurately

(because they match the expected meaning) than unrelated language switches. Instead, in Experiment 2 we found that related and unrelated language switches were detected equally quickly and accurately. Thus, our results suggest that form monitoring completed first and operated autonomously, and meaning monitoring completed only after form monitoring.

We argued that, in an entirely comprehension-based internal monitor such as the perceptual-loop monitor, unintended language switches could only be detected by at least partial processing of form (because form is encountered first, and, according to the theory, the monitor does not have access to the results of production-internal intermediate stages of processing). However, our results are consistent with a monitoring component which accesses words' language membership through language tags independently of their form, and which functions autonomously from, and completes more quickly than, meaning monitoring. Such a component would be inconsistent with the Perceptual Loop theory but consistent with production-based monitoring (in which language tags can be accessed directly, without prior access to phonology), thus extending the relevance of our results to this class of theories.

Faster detection of unexpected language switches relative to semantic errors is consistent with recent ERP evidence that bilinguals access language membership information before semantic information when such access is necessary for task performance (Hoversten, Brothers, Swaab, & Traxler, 2015). These authors examined the relative timing of language membership and meaning activation during visual word recognition by recording go/no-go ERP latencies while bilinguals performed simultaneous semantic and language membership classification tasks. In one half of the experiment, language determined whether a response should be given, and, if so, animacy (living or non-living) determined response hand; in the other half of the experiment, animacy determined response execution and language determined response hand. A comparison of ERP onset latencies for the go/no-go decisions based on language membership or semantic category showed that words were differentiated based on language membership approximately 100 ms faster than they were differentiated based on semantic category. The experimental context in Hoversten et al. is similar to the one in the current study in that, in both cases, bilinguals had to monitor language and meaning to determine whether they should give a response (in the current study, button-press responses upon detection of language or meaning mismatches). In this context, both studies suggest that language membership is accessed before meaning.

Interestingly, measuring articulation times for the five words preceding the errors or switches during reading aloud in Experiment 2 revealed a qualitatively different pattern: Articulation was slower before unrelated than before related words. That is, translation equivalent language switches, which mismatch the target language, did not slow down the articulation of preceding words more than semantically-related within-language errors, which belong to the target language. What is more, encountering a target-language word with unexpected meaning was more disruptive than encountering a wrong-language word with the expected meaning. This pattern indexes greater difficulty to integrate into the preceding context words with unexpected meaning than with expected meaning. It thus suggests a greater sensitivity to meaning than to form during bilingual comprehension, and is consistent with the host of evidence for non-selectivity in bilingual comprehension. It also

suggests that the more automatic processes governing comprehension in reading aloud are qualitatively different, and functionally separable, from the more controlled processes of overt error detection; the two different patterns of results are thus not contradictory. This is because the focus of comprehension is on meaning, while the focus of error detection (in both comprehension, and, by assumption, also in production) is on both form and meaning (and meaning is examined more slowly than form).

Taken together, these results have implications for internal (and external) monitoring in bilinguals and are informative about the mechanisms operating in bilingual comprehension.

First, our results suggest that unintended language switches contained in a pre-articulation speech plan would be detected easily by an efficient form-monitoring process in bilinguals (cf. Altarriba et al., 1996; Moreno et al, 2002). Such an efficient form-monitoring process might also operate during monitoring of overt speech (to the extent that the bilinguals in our experiments detected errors while listening to another person, or themselves). Our results thus imply that unintended language switches should be rarer than within-language whole-word lexical substitution errors in bilinguals' spontaneous speech. However, while bilinguals make very few spontaneous unintended switches (Gollan et al., 2011, 2014; Poulisse, 1999), to our knowledge there is no study comparing these rates to the rates of bilingual within-language errors; future work is necessary to evaluate this possibility.

Second, our results inform perceptual-loop internal monitoring (examining planned utterances for errors through the comprehension system) in monolinguals as well as bilinguals. Previous results are inconclusive about the perceptual loop monitor's sensitivity to form and meaning. For example, Slevc and Ferreira (2006) showed that halting a picture description after detecting a mismatch with a comprehended word (also done by the monitor, according to the theory) was sensitive to that word's phonological similarity with the picture name, but not to meaning similarity. It remains inconclusive, however, how the monitor is able to detect meaning mismatches with intended words (which it presumably has to do too). Our results suggest that the perceptual loop monitor might be sensitive to both phonological form and meaning, but that the meaning-monitoring process might be slower and more error-prone. That is, the fast and accurate form-monitoring process would complete first, and the slower and more error-prone meaning-monitoring process would complete second. The discrepancy between our results and those of Slevc and Ferreira's might be because their participants responded very quickly (under 600 ms in all conditions, compared to 700-800 ms in our Experiment 1) and there was not enough time for the slower meaning-monitoring process to complete.

Note that our conclusions thus far hinge on the assumption that within-language semantic errors and language switches have an equal probability of occurring in a pre-articulation speech plan. That is, in our experimental materials, switches and errors occurred equally often. However, error detection patterns might change if one type of error has a higher probability of occurrence than another. But there seems to be no theoretical reason to assume that this should be the case in natural speech. Specifically, we are not familiar with any evidence suggesting that failures of lexical selection (responsible for within-language semantic errors) should be more frequent than failures of bilingual language control

(responsible for unintentional language switches), or vice versa. Also, in our experimental materials, the rate of occurrence of language switches was 1.5%, possibly higher than the rates of overt production of wrong-language intrusions in bilingual spontaneous speech (<1%; Gollan, Sandoval, Salmon, 2011; Poulisse, 1999). Note, however, that the rates of errors produced overtly are not informative about the rates of error detection *prior to* production (which go unnoticed in overt speech); to our knowledge, the latter are as yet unknown. In other words, the probabilities of planning errors and monitoring success are not necessarily linked. At least in principle, it could be that errors which rarely occur in the pre-articulatory speech plan are easier to detect by the internal monitor (because they stand out), or harder (because the monitor is not attuned to their detection). In any case, a lower rate of wrong-language intrusions in the pre-articulatory speech plan than the rate assumed here does not imply that the pattern of error detection would be different than the one we report (as long as intrusions occur as often as within-language semantic errors).

Beyond implications for the perceptual loop theory, our results provide further support for non-selectivity in bilingual language comprehension (e.g., Libben, et al., 2011; van Assche et al., 2009; 2011; Lagrou et al., 2013). We found that early processing in comprehension (during the eye-voice span in reading aloud) was slowed down more by words drastically mismatching in meaning (e.g., *table*, or *silla* [Sp. chair]) than by translation equivalents which match in meaning but mismatch in language (e.g., *bolsa* [Sp. bag], in the sentence *...she carried too many books in her ...*). Indeed, it is likely that translation equivalents would have disrupted processing more than the actual intended completions (e.g. *bag*), because the latter match both the intended meaning and the intended language perfectly (but were never presented in our experiments; see Altarriba et al., 1996). But what our results show is that the possible processing disruption from words in the unexpected language was less than the disruption by words with unexpected meaning, suggesting that early processing in bilingual comprehension is relatively insensitive to language membership.

Our results, together with Altarriba et al.'s (1996), present a picture of the processes governing bilingual comprehension. In semantically constraining contexts, bilingual comprehenders seem to have early expectations about both meaning and language membership, and violations of either of these expectations disrupt processing (Altarriba et al.). However, violations of expected meaning are much more disruptive than violations of expected language membership. Despite this, bilinguals are aware of unexpected language switches and can detect them quickly and efficiently when necessary (e.g., to alert a bilingual conversational partner who spontaneously code-switches in the presence of a third, monolingual, conversational partner, that the monolingual will not understand the code switch).

In conclusion, we showed that bilinguals detected translation equivalent language switches more quickly and accurately than within-language semantic errors, during both listening and reading aloud. We take these results to suggest that the same might be true for a bilingual internal monitor, which uses sensitivity to form to detect unwanted language switches. More broadly, they suggest that form errors might be detected more quickly and reliably than

meaning errors during both internal and external monitoring in bilinguals as well as monolinguals.

Acknowledgements

This research was supported by an R01 from NICHD (HD050287) and an R01 from NIDCD (DC011492) awarded to Tamar H. Gollan, and by an R01 from NICHD (HD051030) awarded to Victor S. Ferreira. Heartfelt thanks go to Dan Kleinman for programming the experiments reported herein, to Reina Mizrahi for recording and re-recording the stories until they matched our criteria, to Mayra Murillo, Mario Attie and Katia Anguiano, for testing all participants, and to Mayra Murillo, Tiffany Ho and Daphne Liu, for spending countless hours extracting the reading-aloud time-stamps in Experiment 2. The authors declare no conflict of interest.

Appendix A: Experimental materials

The 16 short stories were created by the first author and checked by the last author. The Spanish translations were checked by two native speakers of Mexican Spanish. The words provided inside each story here correspond to Version1 (there were four versions, each presented to different participants). The words in all four versions for all critical sentences are provided below each story, together with English translations of Spanish words. The errors and switches inside stories are bolded here for easy recognition, but, in the actual experiment, they were visually indistinguishable from the rest of the text. Each two stories containing critical sentences ending with the same intended completions are provided here next to one another (stories were presented to participants in a different random order).

Story 1

Mary likes to remember the time when she was a girl. In the big garden they had, she used to play all day long with her sister and her **hermano**¹. The children liked to play many different games. For example, one of them would chase the others around. It was completely safe to let the children play in the garden because it was enclosed by a **meat**². Mary also remembered that, later on, at school, she liked a boy in her class. One day after classes had ended, the boy gave her a red rose and declared his **brazo**³. Otherwise, Mary didn't like going to school. She found most of the classes boring. And her shoulders always hurt because she always carried too many books in her **wallet**⁴. But she always remembered the day she got a rose from the boy she liked. Mary and the boy sat together in the empty classroom for several hours without speaking a single **wall**⁵. She felt like she wanted to say many things but didn't know how to say them. The only thing she could do was smell the flower so she frequently brought it to her **nariz**⁶. She vaguely remembered seeing a few people outside on the street. Most people were walking quickly and not even looking at the school, only one man was standing still by a lamp post. The man slowly lit a cigarette, breathed in and puffed out a lot of **fire**⁷. Finally, Mary got out of the classroom and ran all the way to where she lived. When she arrived, she felt so emotional that she decided to steal some alcohol from her parents. Quietly, she took out a bottle and poured herself some **dedo**⁸. Even though she didn't like going to school, now she really missed those times.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	hermano [brother]	cuchillo [knife]	husband	glass
2	meat	window	leche [milk]	cerca [fence]
3	brazo [arm]	amor [love]	tooth	hate
4	wallet	table	bolsa [bag]	silla [chair]
5	wall	question	puerta [door]	palabra [word]
6	nariz [nose]	mentira [lie]	mouth	truth
7	fire	hat	humo [smoke]	zapato [shoe]
8	dedo [finger]	vino [wine]	eye	cake

Story 9

The Johnsons had two children, Veronica and Richard. Richard always seemed more clever than his sister, and Veronica grew up envying her **cuchillo**¹. Once, she was so mad that she broke all his toys in a rage. Another time, she locked all the entrances and left Richard alone in the yard surrounded by the **window**². When the children grew up, however, their relationship changed. Neither of them married, so the two kept on living together. Veronica became very affectionate - she took care of Richard and cooked for him with **amor**³. The two never did anything together, but once they decided to go on a journey. Veronica always packed light - all she wanted to take with her fit in a **table**⁴. Richard, unlike Veronica, always seemed to need a lot of stuff and had two suitcases. The night before the journey, Richard asked Veronica if she hated him as a child. Veronica assured him she didn't and gave him her **question**⁵. It turned out Richard wanted to apologize for all those years ago, and had brought her beautiful flowers. It was a pity she had a terrible cold and couldn't smell anything because she had a stuffed **mentira**⁶. Talking in the living room, they suddenly saw big flames coming from the kitchen. Veronica started coughing and couldn't breathe because her lungs were filled with **hat**⁷. They called the fire department, but a long time passed before the firemen came. Veronica and Richard were safe, but everything else burned down. They were both in such a shock that they were glad when the neighbors brought them a sip of **vino**⁸. Luckily, their neighbors were nice people and were happy to help out. They gave Veronica and Richard food and clothes and helped them find another home.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	cuchillo [knife]	hermano [brother]	glass	husband
2	window	meat	cerca [fence]	leche [milk]
3	amor [love]	brazo [arm]	hate	tooth
4	table	wallet	silla [chair]	bolsa [bag]
5	question	wall	palabra [word]	puerta [door]
6	mentira [lie]	nariz [nose]	truth	mouth

Critical sentence	Version 1	Version 2	Version 3	Version 4
7	hat	fire	zapato [shoe]	humo [smoke]
8	vino [wine]	dedo [finger]	cake	eye

Story 2

It was the nicest birthday celebration in Alexandra's life. She had a lot of fun trying to blow all forty candles on the **wine**¹. She didn't manage at first so she had to try a second time. Since it was a beautiful summer day, Alexandra was having her birthday celebration outside in the garden. The sun was shining brightly in the sky and there wasn't a single **nube**². It was lucky that the weather was so good, because people could swim in Alexandra's pool. Later on, Alexandra's friends recited some of her poems and read some of her stories. There was a man who did not understand her stories, so at the end of each story he always asked at least one **pregunta**³. One story was about an old and lonely woman. In her solitude, the old woman sewed all the time so she constantly needed needles and some **hilo**⁴. After 3 hours, Alexandra suddenly realized she hadn't eaten anything all this time. She put some fruit on a plate and put a strawberry in her **lie**⁵. Then Alexandra looked again at the man who always asked questions. He hadn't shaved and was dressed in shabby clothes. The only reason his pants did not fall down was that they were kept around his waist by an old **dog**⁶. Alexandra realized it had belonged to her grandfather. Quite out of place, the man wanted to offer her money, so he reached down in his pocket for his **mesa**⁷. The man actually reminded Alexandra of her grandfather, who tormented her as a kid. It was amazing that they looked so much alike. Alexandra disliked her grandfather so much that she was suddenly overcome by an intense feeling of **love**⁸. Alexandra didn't like to be reminded of her childhood.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	wine	finger	pastel [cake]	ojo [eye]
2	nube [cloud]	corbata [tie]	moon	shirt
3	pregunta [question]	muro [wall]	word	door
4	hilo [thread]	cebolla [onion]	chain	orange
5	lie	nose	verdad [truth]	boca [mouth]
6	dog	suit	caballo [horse]	cinto [belt]
7	mesa [table]	cartera [wallet]	chair	bag
8	love	arm	odio [hate]	diente [tooth]

Story 10

One day, Sarah was bored and decided to try something new. She bought chocolate, dough, eggs and nuts because she wanted to bake a delicious **finger**¹. Her friends liked it so much that she decided to open a bakery. Although it didn't work, Sarah didn't give up - she opened a hotel. The hotel sign had a piece of blue sky and a painted sun coming out of a stormy

corbata². The sign read, "Sarah's Hotel - rest well". Sarah did not want to fail again and was trying really hard to be successful. Sometimes, clients were demanding, but Sarah always had an answer to every **muro**³. However, she needed to save money in order to make a profit, so she repaired things herself. For example, she sewed torn sheets herself so she only needed to buy needles and **cebolla**⁴. Sarah always wanted the best quality for her hotel. Once, when she ordered new furniture for the hotel, the beds were not what she wanted. However, she didn't want to argue so she decided not to open her **nose**⁵. One evening, a hotel guest arrived that Sarah didn't like. Around his waist the man was wearing what looked like a **suit**⁶. But it seemed that there were things attached to it hidden under his jacket. The man wanted a room for only one night. The next morning, he pretended he wanted to pay for the room so he reached for an inside pocket in his jacket to take out his **cartera**⁷. Instead, he took out a gun, pointed it at another hotel guest and pulled the trigger. The crime was apparently motivated by prejudice and **arm**⁸. Luckily, the shot missed and the innocent hotel guest escaped unhurt.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	finger	wine	ojo [eye]	pastel [cake]
2	corbata [tie]	nube [cloud]	shirt	moon
3	muro [wall]	pregunta [question]	door	word
4	cebolla [onion]	hilo [thread]	orange	chain
5	nose	lie	boca [mouth]	verdad [truth]
6	suit	dog	cinto [belt]	caballo [horse]
7	cartera [wallet]	mesa [table]	bag	chair
8	arm	love	diente [tooth]	odio [hate]

Story 3

Angela and John liked nature, so they decided to spend the Sunday outdoors. They left home without much preparation. They went to a beautiful valley, where all day long they swam in the **hair**¹. It was dark before they knew it, and they couldn't find their way back to the car. Luckily, they saw a cottage but before they could knock somebody opened the **palabra**². It was a man, who asked them if they needed anything. Both Angela and John were so tired that they couldn't speak. Angela saw a pitcher of water inside and just pointed to it with her **cake**³. The man gave them food, and invited them to spend the night in the cottage. The exhausted couple was really grateful. John suddenly realized he couldn't walk - all along there had been a little stone in his **zapato**⁴. The man told them the cottage once belonged to the king of Spain. Both of them knew this was not possible - all of it was obviously just a **truth**⁵. The cottage was actually a farm, and there were a lot of animals. The next morning, Angela asked if she could ride the **cinto**⁶. Angela and John became friends with the man, and spent a whole week at the farm. They went for long walks in the fresh air. The man had several cows too, so every day at breakfast they drank a lot of **cerca**⁷. One day, they found a wounded deer, and the man took it to his farm. He just couldn't see the deer suffer - he had a kind **corazón**⁸. At the end of the week, Angela and John were sad to go back to the city.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	hair	bridge	cuerpo [body]	rio [river]
2	palabra [word]	puerta [door]	question	wall
3	cake	eye	vino [wine]	dedo [finger]
4	zapato [shoe]	humo [smoke]	hat	fire
5	truth	mouth	mentira [lie]	nariz [nose]
6	cinto [belt]	caballo [horse]	suit	dog
7	cerca [fence]	leche [milk]	window	meat
8	corazón [heart]	avión [plane]	brain	ship

Story 11

The gnomes were little creatures that had lived in the forest for a long time. During the day, they gathered fruit and wood, and in the evening they sat together and sang. At night, they swam in the fastflowing waters of the **bridge**¹. One young gnome was bored with this quiet life and decided to leave. He went to the nearby castle and wanted to walk in, but the guards quickly closed the **puerta**². The young gnome hoped to marry the princess who lived in the castle. When he finally got into the castle, he found the princess and quickly slipped a ring onto her **eye**³. The princess had a guardian fairy who wanted to punish the gnome. She spread the stairs with sticky jam, so the gnome's feet got stuck and he had to leave behind one **humo**⁴. The princess kept the ring from the gnome but was a little afraid it could bring her harm. She went to an old woman for advice, but the old woman was a witch and told her the ring was harmless. In fact, the ring was poisonous and the witch had told a **mouth**⁵. The princess started wearing the ring and soon became very ill. The news that the princess was dying spread quickly, and a handsome prince set off to save her. After avoiding many dangers, the prince arrived, riding a **caballo**⁶. The princess had gotten very thin because she couldn't eat. Following the advice of a good fairy, the prince found a magic goat and told the princess to drink some of its **leche**⁷. But this did not help, and soon the princess lost consciousness. Everybody thought she was dead, because even from up close they couldn't hear the beating of her **avión**⁸. But the prince bent down and kissed her and she came back to life. They got married and lived happily ever after.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	bridge	hair	río [river]	cuerpo [body]
2	puerta [door]	palabra [word]	wall	question
3	eye	cake	dedo [finger]	vino [wine]
4	humo [smoke]	zapato [shoe]	fire	hat
5	mouth	truth	nariz [nose]	mentira [lie]
6	caballo [horse]	cinto [belt]	dog	suit

Critical sentence	Version 1	Version 2	Version 3	Version 4
7	leche [milk]	cerca [fence]	meat	window
8	avión [plane]	corazón [heart]	ship	brain

Story 4

Playing in the yard, the kids started throwing sand at each other. Suddenly one of them cried out that he couldn't see with one **ojo**¹. Luckily there was nothing serious, and soon the family sat down to have dinner. One of the children made their mother angry because he was rocking in his **silla**². The mother served the tasty meal and everybody started eating. There was loud barking coming from the neighbor's place - it was their **horse**³. The kids were excited and were chatting happily. They started making up stories, but their mother got angry again. She told them that they should always be honest and say nothing but the **boca**⁴. The mother always taught her children good manners. She told them that they shouldn't speak while eating. She also told them that, when you entered a Christian temple, you should take off your **smoke**⁵. The mother sometimes helped her children with their school homework. Helping with science homework, she once explained that the skull is made of thick bone and surrounds the **plane**⁶. The mother also had a full time job. Her job was really stressful so she was very busy. Getting ready in the morning, she didn't even have time to look at her reflection in the **picture**⁷. Her work place was also quite far away - she needed more than an hour to get there. One day, she learned that somebody had tried to commit suicide near where they lived. To take his life, the man had tried to jump from a **punte**⁸. This upset the mother so much that she couldn't think of anything else the whole day.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	ojo [eye]	pastel [cake]	finger	wine
2	silla [chair]	bolsa [bag]	table	wallet
3	horse	belt	perro [dog]	traje [suit]
4	boca [mouth]	verdad [truth]	nose	lie
5	smoke	shoe	fuego [fire]	sombrero [hat]
6	plane	heart	barco [ship]	cerebro [brain]
7	picture	newspaper	espejo [mirror]	libro [book]
8	punte [bridge]	cabello [hair]	river	body

Story 12

Patricia was sitting in her favorite coffee shop and enjoying a cup of hot coffee. She suddenly noticed a man was observing her out of the corner of his **pastel**¹. After some time, the man got up and asked Patricia to join him. The man was very polite and seemed very well-educated because he told her a lot of interesting things. But Patricia did not want to seem too interested, so she moved away, leaning back in her **bolsa**². The man told her

everyday things about his life. For example, he said that he went out every evening to walk his **belt**³. He also told her some things which Patricia found strange. He admitted he sometimes heard the voice of his dead mother speaking to him. But Patricia could tell he was honest - he seemed to only be telling the **verdad**⁴. At some point he mentioned his grandfather and what a gentleman he had been. In the presence of a lady, his grandfather had always taken off his **shoe**⁵. The grandfather had also had a lot of character. He had easily gotten angry at people and called them stupid. He would say that inside their skull there seemed to be no **heart**⁶. He had also been very strict with his grandchildren. At some point, Patricia wanted to put some more make-up on, so she went to the bathroom where she could look at herself in the **newspaper**⁷. Soon afterwards, Patricia promised to meet the man again and left. She could walk home, as she only had to cross a big freeway using a **cabello**⁸. Patricia kept thinking about the man from the coffee shop all the way home. But she was not certain any more if she wanted to meet him again.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	pastel [cake]	ojo [eye]	wine	finger
2	bolsa [bag]	silla [chair]	wallet	table
3	belt	horse	traje [suit]	perro [dog]
4	verdad [truth]	boca [mouth]	lie	nose
5	shoe	smoke	sombrero [hat]	fuego [fire]
6	heart	plane	cerebro [brain]	barco [ship]
7	newspaper	picture	libro [book]	espejo [mirror]
8	cabello [hair]	punte [bridge]	body	river

Story 5

Sandra and Peter were both very excited about their wedding. They spent a lot of time planning it. Since both of them were devoted Christians, they wanted to get married in **mano**¹. It was actually going to be a very small wedding. They had only invited their closest friends and relatives. At the reception, everybody would sit around a single **bag**². Everybody would be very happy for Sandra, and her dad would make a toast. Peter asked if she wanted to have live music, and she agreed by nodding her **nieve**³. There were so many things that needed to be done in preparation for the wedding. Sandra wanted to put up a portrait of her and Peter - it was a beautiful **periódico**⁴. Then, she had to send out all the invitations. The invitations were beautiful cards with a romantic picture of a night sky lit up by the **tie**⁵. Deciding what food to serve was another issue, because Sandra wanted everything to be home-made. Her mother helped cook and both of them cried while cutting an **orange**⁶. Maybe they cried a little because of the wedding too. While they were cooking, Sandra looked outside and saw a little child playing. She had an impulse to wave at him, so she stretched out her **hate**⁷. The child happily waved back. The big day finally came and Peter and Sandra started getting ready. Peter put on the pants and jacket of his **traje**⁸. Sandra put on her wedding dress, and everybody thought she was the most beautiful bride they had ever seen.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	mano [hand]	iglesia [church]	leg	house
2	bag	chair	cartera [wallet]	mesa [table]
3	nieve [snow]	cabeza [head]	rain	face
4	periódico [newspaper]	cuadro [picture]	book	mirror
5	tie	cloud	camisa [shirt]	luna [moon]
6	orange	chain	cebolla [onion]	hilo [thread]
7	hate	tooth	amor [love]	brazo [arm]
8	traje [suit]	perro [dog]	belt	horse

Story 13

Travelling around the world, Antonio found himself in a small village in Europe. On the main square, there was a building with a cross on top and a bell tower that he thought must be the **iglesia**¹. Antonio felt quite lost in the foreign village, but suddenly a man went up to him and started gesturing excitedly. Antonio realized the man was inviting him to his place for dinner, and gratefully accepted. When they arrived, there were a lot of tasty-looking dishes and the man's family was sitting around the **chair**². Antonio was overcome with emotion because these complete strangers were so nice to him. He almost broke down thinking how rewarding it was to travel around the world and get to know such nice people. That night a lot of thoughts passed through his **cabeza**³. The man's home was decorated modestly but with good taste. A painted landscape was hanging in the living room; Antonio thought it was such a nice **cuadro**⁴. The family had two children, a boy and a girl. Before bed, the mother read them an educational story about space travel. As far as Antonio could tell, it was about the fact that in 1969, the first man arrived on the **cloud**⁵. When the kids were in bed, Antonio got a strange foreign fruit for dessert. The fruit was white and crunchy and had an intense hot taste like garlic or **chain**⁶. The next morning, the man showed Antonio where the fruit came from. The fruit was hanging high in a tree, so Antonio had to reach out with his **tooth**⁷. In the evening, they went to the village square and the man was wearing his traditional costume. There were colorful decorations on both the jacket and pants of his **perro**⁸. The next day, Antonio had to be on his way, but later he often thought about his stay in the village. The warmth of these people and their hospitality stayed with him for the rest of his life.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	iglesia [church]	mano [hand]	house	leg
2	chair	bag	mesa [table]	cartera [wallet]
3	cabeza [head]	nieve [snow]	face	rain
4	cuadro [picture]	periódico [newspaper]	mirror	book

Critical sentence	Version 1	Version 2	Version 3	Version 4
5	cloud	tie	luna [moon]	camisa [shirt]
6	chain	orange	hilo [thread]	cebolla [onion]
7	tooth	hate	brazo [arm]	amor [love]
8	perro [dog]	traje [suit]	horse	belt

Story 6

Turning 80, Bill thought that he had had a happy life even though he was never rich. He was a carpenter, so the rooms were always full of sawdust. To let some fresh air inside, his wife often opened the **ventana**¹. Bill and his wife lived in California, so the weather was nice year round. For years, they had lived in a small apartment, but finally they saved enough money and could buy their own **church**². Their children came to visit twice a year. Bill's wife had been married before but she never saw her ex **knife**³. Both Bill and his wife liked to have a healthy life. Every morning, they had a cup of juice squeezed from a fresh **cadena**⁴. For lunch and dinner, they had vegetables from the garden. But once, Bill could not eat for a week because he had a strong pain in his cheek and could not chew. Reluctantly, he realized he had to go to the dentist and have him pull out one **diente**⁵. They fell on hard times when carpenters began to be replaced by factories. One by one, Bill's clients left him because the products made in factories were cheaper. Bill had to look for another job, so he started going through the classified ads in the **mirror**⁶. Bill also remembered the day their house almost burnt down. One of their absent-minded nephews didn't put out his cigarette and the curtains caught **sombrero**⁷. Bill didn't make much of it but his wife was really scared. Bill could never forget the terrified expression on her **head**⁸. He quickly poured water on the flames and put them out. Despite the bad times, Bill was content with the way he had lived his life.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	ventana [window]	carne [meat]	fence	milk
2	church	hand	casa [house]	pierna [leg]
3	knife	brother	vaso [glass]	esposo [husband]
4	cadena [chain]	naranja [orange]	thread	onion
5	diente [tooth]	odio [hate]	arm	love
6	mirror	book	cuadro [picture]	periódico [newspaper]
7	sombrero [hat]	fuego [fire]	shoe	smoke
8	head	snow	cara [face]	lluvia [rain]

Story 14

Rita woke up and thought how much she enjoyed her life. It was a beautiful day and sunshine was coming through the **carne**¹. It was spring, the trees were in bloom, and birds

were singing outside. For years, Rita had shared an apartment with other people but now she lived in her own **hand**². It was nice to finally have a place she could call her own. Rita had been married for several years now but had no children. In fact, she really wanted to, so she decided to bring up the topic of children to her **brother**³. It was getting late, so Rita got out of bed and went to the kitchen to make breakfast. She looked in the fruit bowl and decided that the lemons would be too sour, so she squeezed the **naranja**⁴. She then made eggs and toast, and sat down to eat. Accidentally, she bit too hard on her fork and chipped her **odio**⁵. Rita's husband was quite wealthy, so she didn't have to work. She usually spent the day around the house, except for shopping and a few classes she was taking. Every morning, Rita woke up and waited in bed until the mailman brought the **book**⁶. She took it from the mailbox and went out in the garden to read it and drink a cup of coffee. In the evenings, when it got cold, she brought some wood in and started a **fuego**⁷. Rita's husband usually worked late, so she read or watched TV by herself until she fell asleep. Once, Rita saw somebody on TV that looked really familiar. She didn't remember the name but knew she had definitely seen his **snow**⁸. She thought and thought but could not remember where she had seen this man. It turned out to be one of her former housemates who had become a politician.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	carne [meat]	ventana [window]	milk	fence
2	hand	church	pierna [leg]	casa [house]
3	brother	knife	esposo [husband]	vaso [glass]
4	naranja [orange]	cadena [chain]	onion	thread
5	odio [hate]	diente [tooth]	love	arm
6	book	mirror	periódico [newspaper]	cuadro [picture]
7	fuego [fire]	sombrero [hat]	smoke	shoe
8	snow	head	lluvia [rain]	cara [face]

Story 7

The sea pirates lived by robbing traders of their goods. They prepared for attack every time they saw the sails of a **cerebro**¹. Then, they buried the treasures on a desert island. The pirates kept a parrot and a lion as pets. Every evening, the lion ate twenty pounds of **milk**². The youngest pirate had escaped from his wealthy family because he wanted freedom and adventure. At home, he had to be formally dressed and was always nearly strangled by the knot of his **shirt**³. One day, he simply sailed away with the pirates. The oldest pirate was a real character. He walked with crutches because he was missing a **pierna**⁴. He often told stories of his years at sea, and of the many dangers he had overcome. He liked to tell of the time he was held prisoner, and every time he moved there was a clinking sound made from the metal of the **onion**⁵. The pirates avoided going too far North but once, in a storm, they lost their way. When the sea calmed, they were among blocks of ice, and could see flakes of **face**⁶. The youngest pirate felt homesick - he often missed the comforts of his home. Life with the pirates was difficult and full of dangers. Even drinking water was not simple - they

had a giant bucket of water but not a single **esposo**⁷. The beds were hard as stone, and there was often not enough food. The captain's wife was also travelling with the pirates. For a dangerous mission, she needed to get disguised like a man, so she needed to hide her long blonde **body**⁸. She knew she risked her life, but she managed to return safely to the ship.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	cerebro [brain]	barco [ship]	heart	plane
2	milk	fence	carne [meat]	ventana [window]
3	shirt	moon	corbata [tie]	nube [cloud]
4	pierna [leg]	casa [house]	hand	church
5	onion	thread	naranja [orange]	cadena [chain]
6	face	rain	cabeza [head]	nieve [snow]
7	esposo [husband]	vaso [glass]	brother	knife
8	body	river	cabello [hair]	punte [bridge]

Story 15

Bea had a very vivid imagination and every day she imagined her life differently. One day, she imagined she lived a hundred years ago, when travelling from Europe to America meant crossing the ocean on a **barco**¹. Another day, she was sad over the loss of an imaginary pet lamb. She was so upset she became a vegetarian in real life, and stopped eating **fence**². On another day, Bea imagined herself in the past again, at the opera. There was something strange wrapped around one man's neck which looked like a scarf, but Bea guessed it must be what we nowadays call a **moon**³. Bea liked to imagine herself in the past a lot, doing a lot of different things. Sometimes she was the first woman to ever get on skis, and everybody thought she was really brave. At that time, skiing was hard because the ski was not firmly attached to your **casa**⁴. Sometimes, Bea went back in time as much as two thousand years. She was the daughter of an Egyptian pharaoh and had a lot of jewels. Her favorite necklace was one with very fine golden rings forming the **thread**⁵. She was the most beautiful woman in ancient Egypt and people from distant lands came to see her. Next, Bea imagined herself on an expedition to the North Pole. So far north, everything was all white and covered with **rain**⁶. Bea often felt fascinated by how different small things can be in different societies, past and present. In our society, to drink water you use a **vaso**⁷. In ancient societies, you could use a horn or even a skull. Bea's mother always said that Bea was daydreaming too much and would be late for school. To please her mother, Bea would quickly take a shower, get dressed and comb her **river**⁸. On the way to school, however, she would start imagining things again.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	barco [ship]	cerebro [brain]	plane	heart

Critical sentence	Version 1	Version 2	Version 3	Version 4
2	fence	milk	ventana [window]	carne [meat]
3	moon	shirt	nube [cloud]	corbata [tie]
4	casa [house]	pierna [leg]	church	hand
5	thread	onion	cadena [chain]	naranja [orange]
6	rain	face	nieve [snow]	cabeza [head]
7	vaso [glass]	esposo [husband]	knife	brother
8	river	body	punte [bridge]	cabello [hair]

Story 8

Paul was a successful businessman, but he tried to live modestly. He had a personal driver, but he preferred to walk to work. Once, he had forgotten his umbrella, so he knew he would get wet when he felt drops of **lluvia**¹. He sometimes met strange people on the street. A man once asked him if he knew that a dictionary was a type of **libro**². Although he was busy, Paul tried to exercise as much as possible. He was very lean, so you could see the muscles on his **río**³. Paul ate in the same local restaurant over and over again. He liked the place, but the servers could be very absent-minded. One time, he had nothing to cut his steak with, so he had to ask for a **glass**⁴. Every once in a while Paul flew home to visit his parents in his private jet. However, he was afraid of heights, so he felt very uncomfortable flying on the small **ship**⁵. Paul worked in a big international company, and was the most successful person in his family. His father had been a construction worker. Over and over again, his father had laid bricks on top of each other to build yet another **door**⁶. Paul had a very important meeting coming up. He had spent long hours in his office preparing for this meeting. When the day came, he wanted to look perfect, so he shaved and carefully buttoned up his white **luna**⁷. Paul did not want to be late for the meeting, so that day he called his driver. Introducing himself to his business partners, he stretched out his **house**⁸. The meeting went well, and Paul went home content.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	lluvia [rain]	cara [face]	snow	head
2	libro [book]	espejo [mirror]	newspaper	picture
3	río [river]	cuerpo [body]	bridge	hair
4	glass	husband	cuchillo [knife]	hermano [brother]
5	ship	brain	avión [plane]	corazón [heart]
6	door	word	muro [wall]	pregunta [question]
7	luna [moon]	camisa [shirt]	cloud	tie
8	house	leg	iglesia [church]	mano [hand]

Story 16

Tom lived in a place where the weather was bad year-round. When he first moved, he didn't mind the bad weather that much. He even occasionally liked to let the water fall on him, walking under the drops of **cara**¹. But he soon got tired of it and started wondering what to do when the weather was bad and he was not at work. He examined his collection of thick novels because he wanted to find a good **espejo**². The one he decided to read was a captivating crime story. A famous woman was killed under mysterious circumstances and they could not find her **cuerpo**³. An experienced detective was investigating the case. He soon found the woman's corpse in an abandoned building. Near the corpse lay a broken handle, and something was sticking out from the woman's back: the blade of a **husband**⁴. The detective already knew who the murderer was and was about to capture him. To save himself, the murderer went to the airport and flew to another country on the next **brain**⁵. The detective followed him and discovered where the murderer was hiding. In despair, the murderer stole a car and tried to escape, but the detective was following him closely. The murderer drove into a narrow tunnel but panicked and crashed against the bricks of the **word**⁶. The police and an ambulance arrived at the crime scene almost immediately and took the murderer out of the car. He looked dead, and there were stains of blood on the white pocket of his **camisa**⁷. The detective flew back home and declared the case closed. Soon, he started working on another case involving a divorce. One evening, the detective was attacked by a man holding a gun in his **leg**⁸. It was the murderer, who was not actually dead and had managed to escape.

Critical sentence	Version 1	Version 2	Version 3	Version 4
1	cara [face]	lluvia [rain]	head	snow
2	espejo [mirror]	libro [book]	picture	newspaper
3	cuerpo [body]	río [river]	hair	bridge
4	husband	glass	hermano [brother]	cuchillo [knife]
5	brain	ship	corazón [heart]	avión [plane]
6	word	door	pregunta [question]	muro [wall]
7	camisa [shirt]	luna [moon]	tie	cloud
8	leg	house	mano [hand]	iglesia [church]

Appendix B: Item list

Item	English	Spanish
1.	brother husband glass knife	hermano esposo vaso cuchillo
2.	meat milk fence window	carne leche cerca ventana
3.	arm tooth love hate	brazo diente amor odio
4.	bag wallet chair table	bolsa cartera silla mesa
5.	wall door word question	muro puerta palabra pregunta
6.	nose mouth lie truth	nariz boca mentira verdad

Item	English	Spanish
7.	fire smoke hat shoe	fuego humo sombrero zapato
8.	wine cake finger eye	vino pastel dedo ojo
9.	tie shirt cloud moon	corbata camisa nube luna
10.	thread chain orange onion	hilo cadena naranja cebolla
11.	dog horse suit belt	perro caballo traje cinto
12.	hair body bridge river	cabello cuerpo puente río
13.	plane ship heart brain	avión barco corazón cerebro
14.	picture mirror newspaper book	cuadro espejo periódico libro
15.	church house hand leg	iglesia casa mano pierna
16.	head face snow rain	cabeza cara nieve lluvia

Appendix C: Comprehension questions

Story 1

1. What games did the children like to play?
2. Did Mary like going to school?
3. What did Mary decide to do when she got home from school?

Story 2

1. How many times did Alexandra try to blow the candles?
2. What did the man offer Alexandra?
3. Of whom did the man remind Alexandra?

Story 3

1. What day did Angela and John decide to spend outdoors?
2. What did Angela see inside the cottage?
3. Who had a broken leg?

Story 4

1. What did the mother always teach her children?
2. How long did the mother need to get to work?
3. What did the mother learn one day?

Story 5

1. Who had Sandra and Peter invited to their wedding?
2. What kind of food did Sandra want for the wedding?
3. What did Sandra see outside while cooking?

Story 6

1. What was Bill's job?
2. Where did Bill and his wife live?
3. What did Bill and his wife eat for lunch and dinner?

Story 7

1. Where did the pirates bury the treasures they robbed?
2. What pets did the pirates have?
3. Who had to go on a dangerous mission once?

Story 8

1. How did Paul usually go to work?
2. Where did Paul usually eat?
3. How did the meeting go?

Story 9

1. What did Veronica break in her rage?
2. How many suitcases did Richard have?
3. How much time passed before the firemen came?

Story 10

1. What did Sarah buy one day, to try something new?
2. How long did the unpleasant hotel guest stay at the hotel?
3. Who was hurt from the incident at the hotel?

Story 11

1. What did the gnomes do in the evenings?
2. What did the princess's guardian fairy spread on the stairs?
3. What was wrong with the ring?

Story 12

1. Where was Patricia sitting?
2. Whose voice did the man hear sometimes?
3. Did Patricia want to see the man again?

Story 13

1. Why was the man gesturing excitedly?
2. How many children did the family have?
3. Where did the family take Antonio in the evening?

Story 14

1. What season was it that day that Rita thought she enjoyed her life?
2. How long had Rita been married?
3. Who was the man Rita saw on TV?

Story 15

1. Why did Bea become a vegetarian?
2. What sport did Bea imagine herself doing that she was the first woman to do?
3. Whose daughter did Bea once imagine herself to be?

Story 16

1. Who was killed?
 2. Who was investigating the case?
 3. Where was the corpse?
-

Appendix D: Additional analyses

We performed additional analyses without the five items containing mass nouns (which rendered those words ungrammatical in the sentences in which they appeared); these analyses thus included 11 items. The pattern of results in these analyses was identical to the one we report, with the following exceptions. Experiment 1: For RTs, the simple effect of relatedness for within-language errors was marginal [$p = .07$]. For button-press misses, the

simple effect of relatedness for language switches was not significant [$p = .27$]. Experiment 2: For articulation times, the main effect of language was not significant [$p = .13$].

We also performed additional analyses of the data from the first eight stories only (in which the experimental conditions were fully counterbalanced), because the sentences in stories 9-16 were significantly lower in cloze probability than the sentences in stories 1-8. These analyses produced an identical pattern of results to the one we report, with the following exceptions. Experiment 1: For RTs, the interaction between language and relatedness was significant [$Estimate = 48.24$, $SE = 24.56$, $z = 1.96$, $p = .05$]. For errors, the simple effect of relatedness for language switches was marginal [$Estimate = 1.78$, $SE = .95$, $z = 1.88$, $p = .06$]. Experiment 2: For articulation times, the effect of language was not significant [$Estimate = -26.31$, $SE = 20.62$, $z = -1.28$, $p = .20$].

We also performed additional analyses to investigate if the pattern of results we report was influenced by participation in another experiment run jointly with Experiment 2 (see Experiment 2, Materials and procedure). For this purpose, we performed LMER analyses with testing order as predictor (in addition to language and relatedness). These analyses indicated that testing order was not a significant predictor for any of the measures, and, with one exception, did not interact with the predictors of interest [all $ps > .20$]. There was a marginal interaction between testing order and language for the button press misses [$Estimate = -.39$, $SE = .23$, $z = -1.67$, $p = .09$], such that within-language errors were missed equally often whether the current experiment was first or second, while language switches were missed 2% less when the current experiment was second. The 2% difference was not significant [$p = .10$].

References

- Altarriba J, Kroll JF, Sholl A, Rayner K. The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory and Cognition*. 1996; 24:477–492. [PubMed: 8757496]
- Baayen, RH. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press; Cambridge: 2008.
- Baars BJ, Motley MT, MacKay DG. Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning & Verbal Behavior*. 1975; 14:382–391.
- Baker L. Comprehension monitoring: Identifying and coping with text confusions. *Journal of Reading Behavior*. 1979; 11:366–374.
- Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*. 2013; 68:255–278.
- Blackmer ER, Mitton JL. Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*. 1991; 39:173–194. [PubMed: 1841032]
- Buswell, GT. *Fundamental reading habits: A study of their development*. University of Chicago Press; Chicago, IL: 1922.
- Coltheart M, Rastle K, Perry C, Langdon R, Ziegler JC. DRC: A dual-route cascaded model of visual word recognition and reading aloud. *Psychological Review*. 2001; 108:204–256. [PubMed: 11212628]
- Cohen J, MacWhinney B, Flatt M, Provost J. PsyScope: An interactive graphical system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavioral Research Methods, Instrumentation and Computation*. 1993; 25:257–271.

- Costa A, Santesteban M. Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*. 2004; 50:491–511.
- Coulson S, King JW, Kutas M. Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*. 1998; 13:21–58.
- Daneman M, Stainton M. The generation effect in reading and proofreading. Is it easier or harder to detect errors in one's own writing? *Reading and Writing*. 1993; 5:297–313.
- Davis CJ. N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*. 2005; 37:65–70. [PubMed: 16097345]
- Declerck M, Lemhöfer K, Grainger J. Bilingual language interference initiates error detection: Evidence from language intrusions. *Bilingualism: Language and Cognition*. 2016 DOI: 10.1017/S1366728916000845.
- Dell GS. A spreading-activation theory of retrieval in sentence production. *Psychological Review*. 1986; 93:283–321. [PubMed: 3749399]
- Dijkstra T, Grainger J, Van Heuven WJB. Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*. 1999; 41:496–518.
- Dijkstra T, Timmermans M, Schriefers H. On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language*. 2000; 42:445–465.
- Donchin E. Surprise!... Surprise? *Psychophysiology*. 1981; 18:493–513. [PubMed: 7280146]
- Duyck W. Translation and associative priming with cross-lingual pseudohomophones: Evidence for nonselective phonological activation in bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2005; 31:1340–1359.
- Duyck W, Diependaele K, Drieghe D, Brysbaert M. The size of the cross-lingual masked phonological priming effect does not depend on second language proficiency. *Experimental Psychology*. 2004; 51:116–124. [PubMed: 15114904]
- Faigley L, Witte S. Analyzing revision. *College Composition and Communication*. 1981; 37:16–53.
- Gollan TH, Sandoval T, Salmon DP. Cross-language intrusion errors in aging bilinguals reveal the link between executive control and language selection. *Psychological Science*. 2011; 22:1155–1164. [PubMed: 21775653]
- Gollan TH, Schotter ER, Gomez J, Murillo M, Rayner K. Multiple levels of bilingual language control: Evidence from language intrusions in reading aloud. *Psychological Science*. 2014; 25:585–595. [PubMed: 24367061]
- Grainger J, Ziegler J. A dual-route approach to orthographic processing. *Frontiers in Psychology*. 2011; 2:54. [PubMed: 21716577]
- Green DW. Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*. 1998; 1:67–81.
- Hacker DJ, Plumb C, Butterfield EC, Quathamer D, Heineken E. Text revision: Detection and correction of errors. *Journal of Educational Psychology*. 1994; 86:65–78.
- Hahne A, Friederici AD. Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*. 2001; 4:123–141.
- Hartsuiker RJ, Kolk HHJ. Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*. 2001; 42:113–157. [PubMed: 11259106]
- Hartsuiker RJ, Corley M, Martensen H. The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related reply to Baars, Motley, and MacKay (1975). *Journal of Memory & Language*. 2005; 52:58–70.
- Hartsuiker RJ, Pickering MJ, De Jong NH. Semantic and phonological context effects in speech error repair. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 2005; 31:921–932.
- Hoversten LJ, Brothers T, Swaab TY, Traxler MJ. Language membership identification precedes semantic success: Suppression during bilingual word recognition. *Journal of Cognitive Neuroscience*. 2015

- Huetting F, Hartsuiker RJ. Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Language and Cognitive Processes*. 2010; 25:347–374.
- Inhoff AW, Solomon M, Radach R, Seymour BA. Temporal dynamics of the eye-voice span and eye movement control during oral reading. *Journal of Cognitive Psychology*. 2011; 23:543–558.
- Jared D, Kroll JF. Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language*. 2001; 44:2–31.
- Jiang N. Morphological insensitivity in second language processing. *Applied Psycholinguistics*. 2004; 25:603–634.
- Kaan E, Swaab TY. Repair, revision and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*. 2003; 15:98–110. [PubMed: 12590846]
- Kolers PA. Reading and talking bilingually. *American Journal of Psychology*. 1966; 79:357–376. [PubMed: 5968473]
- Kutas M, Hillyard SA. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*. 1980b; 207:203–205. [PubMed: 7350657]
- Lackner, JR., Tuller, BH. Role of efference monitoring in the detection of self-produced speech errors. In: Cooper, WE., Walker, ECT., editors. *Sentence processing*. Erlbaum; Hillsdale, NJ: 1979. p. 281–294.
- Lagrou E, Hartsuiker RJ, Duyck W. Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2011; 37:952–965.
- Lagrou E, Hartsuiker RJ, Duyck W. Interlingual lexical competition in a spoken sentence context: Evidence from the visual world paradigm. *Psychonomic Bulletin & Review*. 2013; 20:963–972. [PubMed: 23483527]
- Laver, JDM. Monitoring systems in the neurolinguistic control of speech production. In: Fromkin, VA., editor. *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. Academic Press; New York: 1980.
- Levelt WJM. Monitoring and self-repair in speech. *Cognition*. 1983; 14:41–104. [PubMed: 6685011]
- Levelt, WJM. *Speaking: From intention to articulation*. MIT Press; Cambridge, MA: 1989.
- Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. *Behavioral & Brain Sciences*. 1999; 22:1–75. [PubMed: 11301520]
- Levy BA, Begin J. Proofreading familiar text: Allocating resources to perceptual and conceptual processes. *Memory & Cognition*. 1984; 12:621–632. [PubMed: 6533431]
- MacKay, DG. *The organization of perception and action: A theory for language and other cognitive skills*. Springer-Verlag; New York: 1987.
- Marian V, Spivey M, Hirsch J. Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language*. 2003; 86:70–82. [PubMed: 12821416]
- Marshall RC, Rappaport BZ, Garcia-Bunuel L. Self-monitoring behavior in a case of severe auditory agnosia with aphasia. *Brain & Language*. 1985; 24:297–313. [PubMed: 3978408]
- Marslen-Wilson, WD. Activation competition and frequency in lexical access. In: Altmann, GTM., editor. *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. MIT Press; Cambridge, MA: 1990. p. 148–172.
- McCallum WC, Farmer SF, Pocock PV. The effects of physical and semantic incongruities of auditory event-related potentials. *Electroencephalography & Clinical Neurophysiology: Evoked Potentials*. 1984; 59:477–488. [PubMed: 6209114]
- McCutchen D, Francis M, Kerr S. Revising for meaning: Effects of knowledge and strategy. *Journal of Educational Psychology*. 1997; 89:667–676.
- Moreno EM, Federmeier KD, Kutas M. Switching languages, switching palabras (words): An electrophysiological study of code switching. *Brain and Language*. 2002; 80:188–207. [PubMed: 11827443]

- Motley MT, Camden CT, Baars BJ. Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*. 1982; 21:578–594.
- Ng S, Wicha NY. Meaning first: A case for language-independent access to word meaning in the bilingual brain. *Neuropsychologia*. 2013; 51:850–863. [PubMed: 23376051]
- Nozari N, Dell GS, Schwartz MF. Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*. 2011; 63(1):1–33. [PubMed: 21652015]
- Oakhill J, Hartt J, Samols D. Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*. 2005; 18:657–686.
- Oppenheim GM, Dell GS. Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*. 2008; 106:528–537. [PubMed: 17407776]
- Özdemir R, Roelofs A, Levelt WJM. Perceptual uniqueness point effects in monitoring internal speech. *Cognition*. 2007; 105:457–465. [PubMed: 17156770]
- Pickering MJ, Garrod S. Self-, other-, and joint monitoring using forward models. *Frontiers in Human Neuroscience*. 2014; 8 Article 132.
- Postma A. Detection of errors during speech production: a review of speech monitoring models. *Cognition*. 2000; 77:97–31. [PubMed: 10986364]
- Postma A, Kolk HHJ. The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research*. 1993; 36:472–487. [PubMed: 8331905]
- Postma A, Noordanus C. The production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech*. 1996; 39:375–392.
- Poulisse, N. *Slips of the tongue: Speech errors in first and second language production*. John Benjamins; Amsterdam, The Netherlands: 1999.
- Roussey J, Piolat A. Critical reading effort during text revision. *European Journal of Cognitive Psychology*. 2008; 20:765–792.
- Schwartz AI, Kroll JF. Bilingual lexical activation in sentence context. *Journal of Memory and Language*. 2006; 55:197–212.
- Schwartz AI, Kroll JF, Diaz M. Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processes*. 2007; 22:106–129.
- Severens E, Janssens I, Kühn S, Brass M, Hartsuiker RJ. When the brain tames the tongue: covert editing of inappropriate language. *Psychophysiology*. 2011; 48:1252–1257. [PubMed: 21361966]
- Slevc LR, Ferreira VS. Halting in single-word production: A test of the perceptual loop theory of speech monitoring. *Journal of Memory and Language*. 2006; 54:515–540. [PubMed: 17917694]
- Sommers N. Revision strategies of student writers and experienced adult writers. *College Composition and Communication*. 1980; 31:378–388.
- Titone D, Libben M, Mercier J, Whitford V, Pivneva I. Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1–L2 intermixing. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 2011; 37:1412–1431.
- Tokowicz N, MacWhinney B. Implicit vs. explicit measures of sensitivity to violations in L2 grammar: An event-related potential investigation. *Studies in Second Language Acquisition*. 2005; 27:173–204.
- Van Assche E, Drieghe D, Duyck W, Welvaert M, Hartsuiker RJ. The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*. 2011; 64:88–107.
- Van Assche E, Duyck W, Hartsuiker RJ, Diependaele K. Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychological Science*. 2009; 20:923–927. [PubMed: 19549082]
- Van Hell JG, Dijkstra T. Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*. 2002; 9:780–789. [PubMed: 12613683]

- Vázquez CA, Glucksberg S, Danks JH. Integration of clauses in oral reading: The effects of syntactic and semantic constraints on the eye-voice span. *Reading Research Quarterly*. 1977; 13:174–187.
- Verhoef K, Roelofs A, Chwilla DJ. Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition*. 2009; 110:84–99. [PubMed: 19084830]
- Winograd P, Johnston P. Comprehension monitoring and the error detection paradigm. *Journal of Literacy Research*. 1982; 14:61–76.
- Yudes C, Macizo P, Morales L, Bajo MT. Comprehension and error monitoring in simultaneous interpreters. *Applied Psycholinguistics*. 2013; 34:1039–1057.
- Zorzi M. The connectionist dual process (CDP) approach to modelling reading aloud. *European Journal of Cognitive Psychology*. 2010; 22:836–860.

Highlights

Bilinguals detected unexpected language switches faster than semantic errors.

There are separate form- and meaning internal monitoring processes in bilinguals.

Form monitoring operates first and is independent of meaning monitoring.

Form monitoring might provide an efficient language monitor in bilinguals.

Early on, bilingual comprehension is more sensitive to meaning than to language.

Figure 1.

Predictions for the speed and accuracy of detecting mismatching words in Experiments 1 and 2. A: Pattern of predicted responses if form-monitoring processes precede meaning-monitoring processes; B: Pattern of predicted responses if meaning-monitoring processes precede form-monitoring processes; C: Pattern of predicted responses if form-monitoring and meaning-monitoring processes operate in parallel.

Figure 2a. Response times detecting language switches and within-language errors in listening comprehension. Error bars represent 95% confidence intervals.

Figure 2b. Button-press misses detecting language switches and within-language errors in listening comprehension. Error bars represent 95% confidence intervals.

Figure 3a. Response times detecting language switches and semantic errors while reading aloud. Error bars represent 95% confidence intervals.

Figure 3b.
Button-press misses detecting language switches and semantic errors while reading aloud.
Error bars represent 95% confidence intervals.

Figure 3c.

Reading times for the five words preceding the error word. Error bars represent 95% confidence intervals.

Table 1

Means and standard deviations for all participants' characteristics

Characteristic	Experiment 1 (N = 40)		Experiment 2 (N = 38)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Age	20.8	1.6	20.4	1.5
Sex (% females)	60	–	56	–
English language				
Age of first exposure to English	3.5	2.3	3.2	2.8
% daily English use now	75.9	19.7	81.6	14.2
% daily English use growing up	54.9	12.9	56.1	20.0
English proficiency self-rating	6.6	0.6	6.7	0.5
English MINT	60.9	2.8	–	–
Spanish language				
Age of first exposure to Spanish	0.6	1.0	0.6	1.3
Spanish proficiency self-rating	5.7	0.9	5.9	0.7
	47.7	9.0	–	–

Note: Self-rated proficiency level was obtained by means of a language history questionnaire, and is averaged across self-ratings for four types of language use (speaking, comprehension of spoken speech, reading, writing). The ratings were made on a scale from 1 ("little to no knowledge") to 7 ("like a native speaker").

Table 2

Experimental conditions for the example item (bag/bolsa - wallet/cartera - chair/silla - table/mesa)

Story	Critical sentence	Intended completion (never presented in its sentence)	Conditions			
			(a) Related language switch	(b) Unrelated language switch	(c) Related within- language error	(d) Unrelated within- language error
1	And her shoulders always hurt because she always carried too many books in her ____.	<i>bag</i>	bolsa [Sp. bag]	silla [Sp. chair]	wallet	table
9	Veronica always packed light – all she wanted to take with her fit in a ____.					
2	Quite out of place, the man wanted to offer her money, so he reached down in his pocket for his ____.	<i>wallet</i>	cartera [Sp. wallet]	mesa [Sp. table]	bag	chair
10	The next morning, he pretended he wanted to pay for the room so he reached for an inside pocket in his jacket to take out his ____.					
4	One of the children made their mother angry because he was rocking in his ____.	<i>chair</i>	silla [Sp. chair]	bolsa [Sp. bag]	table	wallet
12	But Patricia did not want to seem too interested, so she moved away, leaning back in her ____.					
5	At the reception, everybody would sit around a single ____.	<i>table</i>	mesa [Sp. table]	cartera [Sp. wallet]	chair	bag
13	When they arrived, there were a lot of tasty-looking dishes and the man's family was sitting around the ____.					

Table 3
Distribution of an item (bag/bolsa-wallet/cartera-chair/silla-table/mesa) across stories and experiment versions

	Story 1	Story 9	Story 2	Story 10	Story 4	Story 12	Story 5	Story 13
Intended	bag	bag	wallet	wallet	chair	chair	table	table
Version 1								
Related language switch	bolsa							mesa
Unrelated language switch		silla					cartera	
Related within-language error				bag	table			
Unrelated within-language error			chair			wallet		
Version 2								
Related language switch		bolsa					mesa	
Unrelated language switch	silla					table		cartera
Related within-language error			bag					
Unrelated within-language error				chair	wallet			
Version 3								
Related language switch				cartera	silla			
Unrelated language switch			mesa			bolsa		
Related within-language error	wallet						bag	chair
Unrelated within-language error		table						
Version 4								
Related language switch			cartera			silla		
Unrelated language switch				mesa	bolsa			
Related within-language error		wallet					chair	
Unrelated within-language error	table							bag

Note: bolsa – (Sp.) bag; cartera – (Sp.) wallet; mesa – (Sp.) table; silla – (Sp.) chair. The distribution of items across stories was random.

Table 4

LMER results for Experiment 1

Model	Predictors	Estimate	SE	<i>z or t</i>	<i>p</i>
RT					
Main model	Intercept	752.58	18.68	40.29	< .001
	Language	75.97	13.87	5.48	< .001
	Relatedness	22.13	10.14	2.18	.03
	Language * relatedness ⁴	27.72	17.19	1.61	.11
Simple effects of relatedness	Language switches	8.27	8.07	1.03	.31
	Within-language errors	35.99	16.98	2.12	.03
Simple effects of language	Unrelated words	62.12	11.80	5.26	< .001
	Related words	89.83	19.82	4.53	< .001
Button-press misses					
Main model	Intercept	3.32	.15	22.44	< .001
	Language	3.36	.28	11.85	< .001
	Relatedness	1.23	.35	3.58	< .001
	Language * relatedness	.10	.64	.16	.87
Simple effects of relatedness	Language switches	1.25	.64	1.97	.05
	Within-language errors	1.19	.18	6.70	< .001
Simple effects of language	Unrelated words	3.38	.45	7.46	< .001
	Related words	3.35	.36	9.45	< .001

⁴This interaction was significant in the analysis of the data from Stories 1-8 only, in which the critical sentences were of significantly higher constraint than the critical sentences.

Table 5

LMER results for Experiment 2

Model	Predictors	Estimate	SE	<i>t or z</i>	<i>p</i>
RT					
Main model	Intercept	196.83	30.80	6.39	< .001
	Language	60.40	19.94	3.03	.003
	Relatedness	48.13	12.64	3.81	< .001
	Language * relatedness	94.48	21.62	4.37	< .001
Simple effects of relatedness	Language switch	.90	10.72	.08	.93
	Within-language error	95.37	20.93	4.56	< .001
Simple effects of language	Unrelated words	20.38	20.49	1.00	.32
	Related words	84.73	25.46	3.33	< .001
Button-press misses					
Main model	Intercept	2.35	.11	20.71	< .001
	Language	1.24	0.20	6.32	< .001
	Relatedness	.79	.17	4.62	< .001
	Language * relatedness	1.28	.25	5.20	< .001
Simple effects of relatedness ^a	Language switch	.15	.25	.61	.55
	Within-language error	1.44	.17	8.46	< .001
Simple effects of language	Unrelated words	.60	.24	2.51	.01
	Related words	1.88	.22	8.41	< .001
Articulation times for 5 words before error word					
Main model	Intercept	1454.34	47.21	30.81	< .001
	Language	32.10	16.55	1.94	.05
	Relatedness	113.95	20.96	5.44	< .001
	Language * relatedness	104.97	32.07	3.27	.001
Simple effects of relatedness	Language switch	166.12	29.16	5.70	< .001
	Within-language error	61.40	20.36	3.02	.003
Simple effects of language	Unrelated words	84.73	25.46	3.33	< .001
	Related words	20.38	20.49	1.00	.32

^aDue to singular convergence with a full random-effects structure, this model did not contain the items random slope for the interaction term.