Hameau, S., Dmowski, U., & Nickels, L. (2023). Factors affecting cross-language activation and language mixing in bilingual aphasia: A case study. *Aphasiology*, *37*(8), 1149-1172.

Factors affecting cross-language activation and language mixing in bilingual aphasia: A case study

Solène Hameau[1,2,3], Urszula Dmowski[3], & Lyndsey Nickels[1,2]
[1]*School of Psychological Sciences, Macquarie University, Sydney, Australia*
[2]*Macquarie University Centre for Reading, Macquarie University, Sydney, Australia*
[3]*Department of Linguistics, Macquarie University, Sydney, Australia*

Solène Hameau
School of Psychological Sciences
Australian Hearing Hub
16, University Avenue
Macquarie University, 2109, NSW
Australia
solene.hameau@mq.edu.au

Factors affecting cross-language activation and language mixing in bilingual aphasia: A case study

Background: Some bilinguals with aphasia tend to mix languages within a single utterance. Two opposing views attribute this to difficulty with either language control or word retrieval.

Aims: This study investigated the influence of factors that increase activation of the non-target language on the occurrence of language mixing errors. This increased activation predicts more language mixing errors if there is a language control issue, but not if they stem from word retrieval difficulties.

Methods and procedures: A picture naming experiment was conducted with a bilingual individual with aphasia who showed language mixing. We investigated the influence of four factors likely to influence activation of representations of the non-target language on response accuracy, response latency and the occurrence of language selection errors: language, language mode, task, and phonological overlap between the target word and its translation equivalent.

Outcomes and results: The increased activation of the non-target language induced by language mode, task and phonological overlap with the translation equivalent did not lead to an increase in language selection errors when compared to correct responses. This is despite the fact that these factors affected accuracy and response latency, in the direction that is expected in unimpaired bilingual performance.

Conclusions: Results were not consistent with a disruption of the cognitive control needed to respond in the intended language. Instead, they highlight that language mixing in this individual, rather than being "pathological", is instead used as a strategy to potentially improve communication when lexical retrieval difficulties occur. Language mixing behaviours in aphasia may not be due to issues of control and have a communicative value that should be recognised.

Keywords: bilingual aphasia, language mixing, language mode, cognates

Introduction and background

Anomia, the difficulty in retrieving words when speaking, is one of the most prominent and challenging features of aphasia resulting in frustration and communication breakdown. Anomia is believed to result from insufficient activation of the target word relative to other words in the lexicon, resulting in errors and omissions (e.g., Dell et al., 1997; Foygel & Dell, 2000).

When trying to retrieve a word in a given language, individuals with aphasia who speak more than one language must also deal with the influence of words from their other language(s). There is consensus that words from both of a bilingual's languages are activated to some degree in speech production (e.g., Costa et al., 1999; De Bot, 1992; Green, 1998; Hoshino & Thierry, 2011; Titone et al., 2011). While in monolingual speakers there is co-activation of words that are related in meaning and in (phonological) form to the target, in bilingual speakers other words are also active: the target's translation equivalent, or words similar in phonological form to the target but in the other language (e.g., Colomé & Miozzo, 2010; Hameau et al., 2021). This added constraint has been suggested to impede bilinguals' word retrieval relative to monolinguals (e.g., Costa et al., 2006; Gollan et al., 2005; 2008; Ivanova & Costa, 2008; Roberts et al., 2002; Sadat et al., 2016).

Inadvertent language mixing in bilinguals with aphasia

In bilinguals with aphasia there can be inadvertent switching between languages (code-switching). Pathological *switching* refers to unwanted cross-language intrusions that happen *between* utterances and pathological *mixing* (also called blending, e.g., Dash & Ansaldo, 2017) to abnormal cross-language intrusions happening *within* a single utterance Fabbro, 2001). In one account, pathological switching and pathological mixing have been hypothesised to result from the same underlying cause: damage to the network responsible for language control (e.g., Green & Abutalebi, 2008). Another account hypothesises different causes: Pathological *switching* is suggested to result from an impairment of the language control system and is often associated with executive control difficulties (e.g., Fabbro et al., 2000; see Mooijman et al., 2021, for a review), but pathological *mixing* is interpreted in the context of lexical retrieval difficulties (e.g., Ansaldo et al., 2010; Goral et al., 2019; Lerman et al., 2019; Muñoz et al., 1999). In some cases, these instances of mixing may in fact only be *abnormal* in that they happen more frequently than expected within the person's language community, as opposed to being qualitatively different (see e.g., Bhat & Chengappa, 2011, for a review). Indeed, for some societies or language communities, and/or in certain pragmatic situations, it is normal to mix languages (e.g., Cheng & Butler, 1989; Grosjean, 1998). Bilinguals may use language mixing to improve the flow of conversation (e.g., Cheng & Butler, 1989), and in bilinguals with aphasia, as a compensation for word retrieval failures, with the aim of increasing communication effectiveness (e.g., Ijalba et al., 2012).

Whatever the cause of language mixing in bilingual aphasia (control, strategy, or both) it demonstrates that, at some point, the target word was more active in the non-target language than in the target language of the interaction. Several factors are argued to influence the state of activation of the target word's translation equivalent, and hence the potential occurrence of mixing.

Factors affecting cross-language activation in bilinguals

Language proficiency and status

The more dominant or proficient a bilingual's non-target language, the more likely it is that this language has a higher state of activation (e.g., in the bilingual Stroop and picture-word interference paradigms: Chen & Ho, 1986; Mägiste, 1984; Tzelgov et al., 1990). Several theoretical accounts (for example, the Revised Hierarchical Model of bilingual memory: e.g., Kroll & Stewart, 1994; Kroll et al., 2010) provide explanations for the difference in the activation status of each language in (unbalanced) bilinguals. For instance, Kroll and colleagues interpret this difference in terms of stronger links from L2 (understood within this framework as the "weaker" language rather than the later acquired language) to L1 (the "stronger" language), as well as there being stronger links between concepts and L1. However, in other theories all words are activated irrespective of language (as in the BIA+ model: Dijkstra & Van Heuven, 2002). In such a theory, words in the "weaker" language may have a lower frequency than words in the stronger language, and therefore less activated and less accessible. Alternatively, when using the "weaker" language greater inhibition may be needed to resolve competition with words from the "stronger" language, leading to a disadvantage (e.g., Green, 1998). Consequently, under all of these accounts, the greater the difference in the strength of the two languages, the more the non-target, stronger language will interfere with lexical retrieval in the weaker, target language.

Language mode

The language mode theory (Grosjean, e.g., 1989, 1998, 2001, 2008) predicts that the strength of activation of the non-target language varies depending on the context of the interaction (e.g., what is/are the language(s) of the conversation partner) and which language is intended by the bilingual speaker. At any given time, bilinguals are on a continuum from a monolingual mode to a bilingual mode. In monolingual mode, when talking to a monolingual speaker of one of their languages, the activation of the bilingual's non-target language is minimal. However, in bilingual and intermediate modes, when communicating with other bilinguals who share both their languages, activation of both languages is enhanced and bilinguals may allow the two languages to mix (e.g., Dewaele, 2001; Grosjean, 1997; Treffers-Daller, 2011). For discussion, see also Green, 2011 and Green & Abutalebi, 2013, about how language mixing patterns induced by the language context might differently influence cognitive control demands. This led Grosjean (1998) to suggest that in people with aphasia who exhibit language mixing behaviours, these behaviours would increase in a context evoking bilingual mode, as compared to monolingual mode. However, as far as we are aware, this prediction has not been directly tested.

Language task

Task demands could also influence the relative activation of a bilingual's languages. For instance, more activation of both languages would be expected in translation and interpreting tasks, in experimental paradigms based on language switching, or in those involving stimuli in the other language, compared to tasks and stimulus materials that involve a single language (e.g., Grosjean, 1998, Kroll & Dijkstra, 2010).

Language similarity

Similarity between the languages of a bilingual is also known to influence cross-language influences. For instance, in the bilingual Stroop paradigm, Brauer (1998) observed more interference between similar than dissimilar language pairs. In a trilingual speaker with aphasia, Goral et al. (2006) found an influence of language similarity on the activation of a non-target language. There was more interference between French and English, than between either of these two languages and Hebrew. Note, however, that age and context of acquisition could also contribute to explaining the observed patterns in this case.

In addition to similarity across languages as a whole, similarity between translation equivalents of a word also influences processing. For example, words that share formal similarity with their translation equivalent (cognates, e.g., English-Spanish translation equivalents *rat*-*rata*) are produced faster and more accurately than those with less similarity (e.g., Costa et al., 2000). Some researchers (e.g., Costa et al., 2000; 2005) have interpreted this cognate facilitation effect as the consequence of interactivity between lexical and post-lexical levels of processing within and across languages of a bilingual. Others have suggested that cognates may act as triggers for code-switching behaviours (e.g., Broersma, 2009; Broersma & De Bot, 2006). In aphasia, cognates can facilitate picture naming performance (e.g., Hameau & Köpke, 2015; Lalor & Kirsner, 2001; Roberts & Deslauriers, 1999), promote transfer of therapy benefits across languages (Kohnert, 2004), or, in contrast, hinder intervention success (Kurland & Falcon, 2011).

Given the multiplicity of factors that may affect cross-language activation and therefore the ease of retrieving the target word, it seems warranted to examine how these factors interact with the retrieval of words in the intended language in bilingual aphasia. Therefore, we systematically investigated how the factors: language, language mode, task, and phonological overlap between translation equivalents; affected the picture naming performance of a (late) bilingual individual with aphasia, who seemed to present with inadvertent language mixing behaviours. If a disruption of inhibitory control is responsible for an increase in language errors in bilingual aphasia (Abutalebi & Green, 2007; Green, 1998; Paradis, 2004), then these factors should increase language selection errors by enhancing activation of the non-target language (Green & Abutalebi, 2008). Alternatively, if the prevalence of language selection errors is not sensitive to these factors, then this type of error is more likely to reflect general word-finding difficulties, and therefore potentially strategic effects. This would be consistent with other accounts of the origin of pathological mixing in bilingual aphasia (e.g., Ansaldo et al., 2010; Goral et al., 2019; Lerman et al., 2019; Muñoz et al., 1999).

Method

This study was approved by Macquarie University Human Research Ethics Committee (ethics number 5201200905).

Participant

The participant, that we call CA, was recruited through a local speech pathology clinic and gave written consent. CA (see Table 1) was a 77-year-old right-handed Spanish-English bilingual male, who migrated to Australia, from a South American country where Spanish is the official language. CA had 15 years of formal education and worked, for most of his career, in a white-collar occupation that involved extensive reading, writing, and communication with the public, in English. CA reported no speech, language, or reading difficulties as a child. Originally left-handed, CA became right-handed upon attending school and being strongly encouraged to write with his right hand. CA experienced a left middle cerebral artery stroke five years prior to the study. The initial CT scan reported focal lesions involving the grey and white matter in the left frontal and parietal lobes, above the left lentiform nucleus and the corona radiata, in the interior limb of the left internal capsule, and in the lateral aspect of the left temporal pole. The initial neurological examination noted the presence of aphasia and right sided facial droop, as well as mild-moderate functional deficit to the right upper limb. CA was briefly reviewed by a speech pathologist while an inpatient and received only 10 language therapy sessions in English after discharge from hospital.

<insert Table 1 about here>

At the time of the study, CA presented with chronic aphasia. He was living alone and was completely independent in his activities of daily living, including managing and driving himself to appointments. He was selected for this investigation as he exhibited frequent language mixing when speaking in English, even when aware that his conversation partner did not understand Spanish. CA taught himself a technique to compensate for this by typing the inappropriate Spanish word into a translation app on his mobile phone, to find the English translation and produce it correctly by reading it aloud.

Background assessments

A series of assessments were conducted with CA (see Appendix A). CA reported being a late Spanish-English bilingual with high pre-stroke proficiency in both languages (see Table 1).

Language dominance is often understood as the balance of proficiency between languages of a bilingual. An alternative, psycholinguistic, definition of language dominance posits that the dominant language is the one that is currently the most easily accessible to the speaker, irrespective of proficiency. In this view, language dominance is also highly dependent on immediate language use (for a discussion, see Köpke and Genevska-Hanke, 2018). As we were interested in word retrieval in bilingual aphasia, this definition of language dominance seemed more suitable. Under this definition, CA was most likely more Spanish-dominant for the nine years before this study, that is, after retiring, and both before and after his stroke. CA's non-linguistic cognitive screening suggested good cognitive control abilities. CA presented with moderate non-fluent aphasia in both languages, with agrammatism. Results of single-word level assessments suggested an impairment in both languages at the level of, or access to, the phonological output lexicon, with some impairment of the phonological output buffer. Based on CA's language history, on his self-report of high pre-morbid proficiency in both languages, and on the background test results, CA was experiencing parallel recovery of his languages at the time of the study.

Study design

The experimental task consisted of an object picture naming task in both English and Spanish, administered in three different ways: 1) monolingual picture naming, administered in either English or Spanish, 2) bilingual picture naming in which CA had to name a picture in one language but with both languages being used during the session, and 3) a "subsequent-retrieval-of-the-word-in-the-other-language" task in which CA had to retrieve a given word in one language immediately after having to retrieve that word from a picture in the other language. Conditions 2) and 3) occurred consecutively, within the same session (see Procedure).

Materials

Coloured picture stimuli with over 85% name agreement (i.e., more than 85% of control participants gave the target name) in Spanish were selected from the Multipic database (Duñabeitia et al., 2018). Homophones and compound words were excluded with assistance from two Spanish speakers of the same Spanish dialect as CA (a trained speech language pathologist and an audiology engineer), leaving a total of 250 target words in each language.

For each target, several item-specific variables were obtained. Subjective visual complexity values (ratings from 1 to 5) were taken from the Multipic database (Duñabeitia et al., 2018): average of English and Spanish visual complexity values, range = 1.17-3.66, average = 2.45, SD = 0.49. Name agreement values were also drawn from the Multipic database (Duñabeitia et al., 2018): Spanish range = 85.56-100, average = 97.92, SD = 3.53; (British) English range = 83.33-100, average = 96.60, SD = 4.33. Word frequency values were taken from the subtitles databases SUBTLEX-UK (van Heuven et al., 2014) and SUBTLEX-ESP (Cuetos et al., 2011). They consisted of the log of the frequency per million of each word (token) in the respective database (1 was added to each value before logging it): Spanish range = 0.07-3.14, average = 1.09, SD = 0.55; English range = 0.14-3.25, average = 1.23, SD = 0.49. Measures of length in phonemes for English targets were taken from the MRC Psycholinguistic Database (Wilson, 1988), and for Spanish words, a phoneme count was conducted on the IPA transcription of each word: Spanish range = 3-11, average = 6.16, SD = 1.82; English range = 1-10, average = 4.40, SD = 1.55. Finally, a measure of phonological distance between translation equivalents was computed for each pair, using the R package AlineR (Downey et al., 2017) based on the ALINE algorithm (Kondrak, 2000). This continuous, objective measure was preferred over a dichotomous, subjective measure of cognate status, as it has the potential to account for more fine-grained effects of the cross-language (dis)similarity between single words. AlineR phonological distance scores are comprised between 0 and 1, with 0 corresponding to identical forms, and 1 to very dissimilar forms: translation pairs range = 0-0.89, average = 0.45, SD = 0.22.

Procedure

A picture naming task was conducted in both English and Spanish using the 250 target items. Pictures were named in each language three times, each corresponding to a different condition or task: monolingual, bilingual, and subsequent-retrieval-in-the-other-language. Monolingual trials were in sessions conducted in only one of CA's languages, and where the task was to name the pictures in that language only. Bilingual and subsequent-retrieval-in-the-other-language trials (hereafter "subsequent-retrieval") occurred during the same session, when responses in both languages were required. In these sessions, upon presentation of a picture, CA had to name that picture in one language ("bilingual" trial response), then immediately in the other language ("subsequent-retrieval" trial response). For example, upon seeing the picture of a mouse, CA would have had to respond *mouse,* then *ratón;* or *ratón,* then *mouse,* depending on the requested order. Then the procedure was repeated with the next target word. Hence, bilingual and subsequent-retrieval trials alternated. In total, the experiment was conducted over six sessions spread over a period of three weeks (see Table 2). Each session approximately lasted one hour, and sessions were separated by at least two days, to limit practice effects.

<insert Table 2 about here>

The experiment started with four monolingual sessions, each targeting half of the items (=125 items), in one single language, alternating language and item set between sessions. During these monolingual sessions a monolingual language mode was encouraged using the following strategies. First, testing was conducted by a native speaker of the language (trained in speech pathology) who either did not know the "other" language (English monolingual trials) or presented himself as a monolingual speaker of the target language (Spanish monolingual trials). Second, a few minutes were taken before the session to chat informally in the language of the session. Third, spoken and written instructions were given in the language of the session. Fourth, a picture of a flag representing the language of the session (i.e., Flag of Australia or of CA's country of origin) was displayed on the table as a further cue to maintain the target language.

The last two sessions were the bilingual/ subsequent-retrieval sessions. In these sessions, strong activation of both languages was facilitated using the following strategies. First, sessions were conducted by a native speaker of the same variety of Spanish as CA (the same speaker who conducted the earlier Spanish monolingual sessions), who was also highly proficient in English. Second, this bilingual examiner engaged in an informal chat at the beginning of the session, in which he mixed Spanish and English in the most "natural" way possible. Third, spoken and written instructions were given in a mix of English and Spanish (for example, written instructions were "Usted verá una serie de imágenes. Say the name of the picture in English first, then in Spanish. Presione espacio para continuar"). Fourth, pictures of an Australian flag and a flag of CA's country of origin were placed on the table, arranged according to the task requirement (i.e., Australian flag placed on the left of the other flag for trials requiring English then Spanish)[1]. In the first of these two bilingual/subsequent-retrieval sessions (i.e., the fifth session), the first 125 items were to be named first in English, then in Spanish, and then the last 125 items were to be named first in Spanish, then in English. For the next session, item half and language order were counterbalanced in a similar fashion, but with the first 125 items first named in Spanish, then in English, and the last 125 items first named in English, then in Spanish..

Stimulus pictures were presented in PowerPoint, on a MacBook laptop computer concurrently with an audible beep. Slides advanced automatically. Each picture was displayed on a single slide for a total of either eight seconds (monolingual trials) or ten seconds (bilingual/subsequent-retrieval trials), after which the following slide was presented automatically with the next picture and concurrent “beep”, and so on. Consistent with Evans, Hula, Quique and Starns (2020), the duration of picture presentation was personally tailored to CA son the basis that he had usually been able to give a response within that cut-off in a pre-study naming trial using an independent set of pictures. The cut-off also sits within the bracket (5 to 10 seconds) which was deemed optimal by these authors. Pictures were presented in a fixed, pseudo-random order. All sessions were audio-recorded with an Olympus WS-852 digital voice recorder.

Response coding

Responses were transcribed and coded for accuracy by the second author and then submitted to two independent assessors for the English and the Spanish responses (either master students in Speech Pathology or a researcher in Audiology) for agreement. Overall, the discrepancy rate overall was low and resolved through consensus, but in case of disagreement, the Spanish-speaking assessors’ transcriptions and coding corrections were used. The final response coding scheme is fully detailed in Appendix B. Response type was coded in terms of accuracy (correct vs incorrect), and incorrect responses were further coded. For the purpose of this study, the only error types considered were “correct response in the other language”, and “omission”. Response latency was obtained manually for correct trials using Audacity (Mazzoni et al., 2005). In monolingual and bilingual trials, the time between the onset of the beep (when the picture was presented) and the onset of the participant’s response was measured. In subsequent-retrieval trials, response latency was the time between the onset of the participant’s previous (correct) response in the other language, and the onset of his response in the target language of the subsequent-retrieval trial. For example, in a trial where the (correct) responses were *conejo, rabbit*, response latency for the response *rabbit* was measured from the onset of *conejo* to that of *rabbit*. Some correct trials were not analysed for latency because of difficulty determining a reliable response time, for example, because the word was produced after a hesitation, as in *curt-curtain* (see Appendix B).

Analyses

Logistic and multiple regressions were computed to investigate the influence of condition, language, and phonological distance between translation equivalents, and their interactions, on accuracy, response time, and likelihood of a language selection error as compared to a correct response, or as compared to an omission. Analyses were run in R (R Core Team, 2014) using the package *car* (Fox & Weisberg, 2019), and post-hoc tests were performed using the package *emmeans* (Lenth, 2021). In all analyses, continuous predictors were standardised. Factors (language, condition) were sum coded. In all analyses, a model-building approach was used, in which the full model including all predictors and interactions was subsequently simplified using model comparison to remove non-significant interactions and/or predictors, resulting in a final model (reported below) that only included significant predictors and interactions. Model-building steps are described in Appendix C.

Accuracy analyses

Of the 1500 trials, 1498 were valid trials (i.e., without an administration error). Of these, responses corresponding to acceptable alternatives (e.g., *lolly* instead of *sweet*, n = 72) and timing/sequence issues (n = 32) (see Appendix B for explanation of response types) were further discarded, leaving a total of 1394 responses considered in accuracy analyses. A logistic regression was run on those trials using accuracy as a dependent variable (correct responses were coded as 1, incorrect responses as 0). Length in phonemes, log word frequency, name agreement, and visual complexity of the intended target word, were used as control predictors. Condition (monolingual, bilingual, subsequent-retrieval), target language, and phonological distance were used as experimental predictors. The significance of a three-way interaction between condition, target language, and phonological distance was tested.

Latency analyses

Multiple regressions were run on latencies for the correct trials. However, because latencies in the subsequent-retrieval condition were differently estimated compared to the other conditions, separate analyses were conducted for a) the monolingual and bilingual conditions (617 trials), and b) the subsequent-retrieval condition/task (269 trials). As for accuracy analyses, length in phonemes, log word frequency, name agreement, and visual complexity of the intended target word were used as control predictors. For the model investigating latency in the monolingual and bilingual condition, condition (monolingual, bilingual), target language, and phonological distance were used as experimental predictors, and the significance of a three-way interaction between condition, target language, and phonological distance was tested. For the model including only response times on the subsequent-retrieval condition, experimental predictors were target language and phonological distance, and a two-way interaction between target language and phonological distance was tested. To approach normal distribution, appropriate transformation of response time values was determined via a BoxCox test (Box & Cox, 1964) using the r package *MASS* (Venables & Ripley, 2002) and was applied to response times in latency analyses.

Language error analyses

Further analyses were carried out to determine what affected the probability of selecting the wrong response language when naming pictures in the three different conditions (monolingual, bilingual, and subsequent-retrieval). To this end, two logistic regressions were performed on a) correct response (coded as 1) vs response corresponding to the target word but in the incorrect language (coded as 0; hereafter "language error"), and b) language error (coded as 1) vs omission (coded as 0), both models using the same predictors and interactions as in the accuracy analysis.

Results

Accuracy analyses

Table 3 below presents total accuracy, and accuracy for each language, condition, and condition within language. Invalid trials, acceptable alternatives, and timing/sequence errors are not included in the total count of responses.

<insert Table 3 about here>

Model results for accuracy are summarised in Table 4. There was no significant effect of condition on accuracy nor any interactions between the experimental condition and any other experimental predictor. Hence, the factor "condition" was not retained in the final accuracy model. A significant effect of language was observed, with Spanish trials being more accurate than English trials overall. There was an effect of phonological distance. Pictures with phonologically similar names in Spanish and English were more often correctly named than pictures whose names were dissimilar. There were no significant interactions involving phonological distance and any of the other experimental predictors (all $p > .05$). Effects of control predictors were all significant and in the expected direction: CA was more likely to correctly name pictures which were visually less complex, which represented words that were more frequent, shorter, and had higher name agreement.

<insert Table 4 about here>

Latency analyses

Raw average response times by condition and language are presented in Table 3 above.

Response times for the Monolingual and Bilingual conditions only

A Box Cox test determined the appropriate transformation as raising response time values to the power of -0.75, an inverse transformation leading to higher transformed latency values corresponding to shorter latencies[2]. There was a significant interaction between condition and language (see Table 4, Figure 1). Overall, responses were slower in the bilingual condition than in the monolingual condition, but this difference was driven by CA's performance in

Spanish, with no difference between response times across conditions for English (estimated marginal means for monolingual vs bilingual conditions: Spanish: Estimate = -0.0007, t.ratio = -4.536, $p < .0001$; English: Estimate = -0.00004, t.ratio = -0.257, $p = .994$). Phonological distance influenced response time, with faster latencies for words with a more similar translation equivalent. There were no significant effects of any control predictors.

<insert Figure 1 about here>

Response times for the subsequent-retrieval *condition only*

Following a Box Cox test response times were raised to the power of -0.2[3]. There was a significant effect of phonological distance, with words that were phonologically more distant to their translation equivalent were produced slower in this subsequent-retrieval condition. There was no effect of language on response time, nor an interaction between language and phonological distance. No control variables reached significance.

Language error analyses

Overall, language errors (correct target name but in the non-target language) were the most frequent error type, representing 29% of errors (see Table 5).

<insert Table 5 about here>

Correct vs language error

For the response comparison correct vs language error, there were significant effects of language and of condition (see Table 6 for a summary of the model results). CA was more likely to produce a language error in English than in Spanish relative to a correct response. While he was no more or less likely to produce a language error in the monolingual compared to the bilingual condition (estimated marginal means for monolingual vs bilingual: Estimate = -0.044, z.ratio = -0.229, $p = .971$), he was less likely to make language errors in the subsequent-retrieval condition compared to the two other conditions (marginal means for bilingual vs subsequent-retrieval: Estimate = -1.369, z.ratio = -4.573, $p < .0001$; monolingual vs subsequent-retrieval: Estimate = -1.324, z.ratio = -4.409, $p < .0001$). There was no effect of phonological distance on the probability of a language error compared to a correct response, and none of the tested interactions involving experimental variables reached significance (all $p > .05$). The control variables frequency and length had a significant effect, with more frequent and shorter words more likely to result in a correct response than in a language error.

<insert Table 6 about here>

Language error vs omission

When language errors were compared to omissions, significant effects of condition and phonological distance were observed (see Table 6 for a summary of the model's results). Words that were more similar to their translation equivalent were more likely to result in a language error than in an omission. Words were less likely to result in a language error than in an omission in the subsequent-retrieval condition compared to the other conditions. There was no significant difference between the estimated marginal means for the monolingual and bilingual conditions (Estimate = -0.263, z.ratio = -0.664, $p = .784$), but significant differences between monolingual and subsequent-retrieval (Estimate = 3.429, z.ratio = 7.959, $p < .0001$), and between bilingual and subsequent-retrieval (Estimate = 3.166, z.ratio = 7.780, $p < .0001$). In addition, there was a significant effect of target frequency, with more frequent words more likely to result in a language error than an omission.

Discussion

The aim of this research was to investigate the origin of inappropriate language mixing errors in bilingual aphasia, through a systematic investigation of the influence of factors likely to increase cross-language activation. We examined four predictors of picture naming behaviour: language, language mode, task, and phonological distance; in CA, a Spanish-English bilingual man with aphasia, and all showed some effect on performance.

Language of the expected response affected CA's performance in the predicted direction: consistent with pre-stroke patterns of proficiency and dominance, performance was either comparable or more accurate in Spanish than in English. Performance was faster in Spanish than in English in the monolingual condition, and CA was also more likely to produce a correct response than a language error when the target language was Spanish than when it was English.

Language mode theory predicted poorer performance in the bilingual condition compared to the monolingual condition. This prediction was only confirmed in the response latency analyses, with slower performance overall in the bilingual than in the monolingual condition. However, the task that we chose to induce a bilingual mode also involves switching. Following the subsequent-retrieval-in-the-other-language response to one picture, CA was required to switch back to the other language for the bilingual trial on the next item. (Forced) language switching, as examined in language-switching paradigms, is known to incur a cost for response latency (e.g., Bonfieni et al., 2019). Consequently, CA's slower response time in the bilingual condition could be due to switching costs.

Crucially, the absence of a difference in the rate of language errors between the monolingual and bilingual conditions shows that a more bilingual language mode (with higher activation of the non-target language from the language context of the task) did not cause more occurrences of language mixing. This is contrary to Grosjean's (1998) speculations on language mixing in Perecman's (1984) study of HB, a bilingual person with aphasia. As these language mixing occurrences had been observed while HB was in bilingual mode, Grosjean alluded to the possibility that HB may have produced fewer language mixing errors in a monolingual context. This prediction did not hold in our experiment.

Nevertheless, our finding still aligns with the principles of language mode theory in the sense that this theory relates to the pragmatic use of language. Indeed, a bilingual individual adapts to the pragmatic constraints (or loosening of constraints, in the case of true bilingual mode) of the communication situation. In our experiment, even though the bilingual condition aimed to generate strong activation of both languages, the pragmatic constraint was still to speak only in the target language, and CA mostly complied with this constraint. A possible caveat restricting our interpretation is that perfect monolingual language mode probably was not (and could not) be achieved for Spanish given that the study took place in an English-speaking country: CA could easily have figured out it was unlikely that the Spanish-speaking experimenter did not know any English, contributing to a less strict monolingual mode (Grosjean, 1998; Hermans et al., 2011). Nevertheless, while we believe this is an important methodological point for bilingual research in general, for our study it is critical that language errors were most frequent in English, not Spanish, and this condition was indeed a “pure” monolingual condition.

If, as has been proposed, interference from the non-target language is greater for the non-dominant language (e.g., Costa et al., 2000), then, under bilingual mode conditions, Spanish should interfere more with English, than English with Spanish. Yet the reverse was the case: the response time difference between the monolingual and the bilingual conditions was larger in Spanish than in English.

However, again, one should examine these results considering that our bilingual condition involved switching. When a bilingual speaker is forced to switch between languages, switching costs are often found to be higher when switching from the non-dominant to the dominant language (e.g., Costa & Santesteban, 2004; Meuter & Allport, 1999; Zheng et al., 2018; but see Christoffels et al., 2007 and Declerck et al., 2012, for symmetrical switching costs in unbalanced bilinguals). Therefore, if bilingual trials are in fact *switch* trials, then CA’s pattern of results is in line with that expected for unimpaired late bilinguals: greater interference in the Spanish bilingual mode (that is, when switching back to Spanish, the dominant language) as compared to the English bilingual mode.

Also in line with language mode theory, we expected accuracy and language errors to be affected by the specific task (picture naming vs subsequent-retrieval-in-the-other-language). Contrary to this prediction, there was no influence of task on general accuracy: CA was no more or less likely to make errors in the monolingual, bilingual, or subsequent-retrieval conditions. However, when looking more specifically at language selection errors, these were more likely in the monolingual and bilingual conditions than in the subsequent-retrieval condition, both when compared to correct responses and when compared to omissions. In other words, CA was more likely to make language selection errors when producing a word from a picture only (i.e., in the monolingual and the bilingual conditions) than when producing it having both the picture and the word that he just produced in the other language. This could be due to several mechanisms. First, in the subsequent-retrieval condition, there is an inherent strong reminder of the target language to be used: in essence, "the other one". This may make language selection errors less likely. Second, provided the semantic representation of the target word was correctly retrieved before the subsequent-retrieval trial (for example, having correctly retrieved the concept of a rabbit and then the word "rabbit" in the first language), CA no longer needed to retrieve semantic information in the subsequent-retrieval task. He could then focus on retrieving the correct word form only, therefore having fewer steps to process and reducing the scope for errors. A related explanation would be that CA completed the subsequent-retrieval task by translating the previously retrieved word. If translation proceeds via word association rather than via concepts as is required in picture naming (for a discussion, see e.g., Kroll & Tokowicz, 2005), such a translation strategy could lead to facilitated processing, and therefore fewer language selection errors in the subsequent-retrieval task. Although this possibility is embedded in the Revised Hierarchical Model (Kroll & Stewart, 1994), it has been challenged (see, e.g., La Heij et al., 1996, or Brysbaert and Duyck, 2010). In addition, this account is usually proposed for L2 (weaker language) to L1 (stronger language) translation only (e.g., Chen & Leung, 1989), rather than in both directions as is the case here. A final possibility is that facilitated retrieval in the subsequent-retrieval condition comes from converging information on the target word form from both semantics and the other translation equivalent via direct links (combined word association and concept-mediated route for translation, as in the Revised Hierarchical Model, Kroll & Stewart, 1994). In sum, although the subsequent-retrieval task implied maximal activation of both languages, this task did not lead to worse performance for CA, as would have been predicted if there was an issue with language control. Instead, CA's performance was better supported in the subsequent-retrieval task than in picture naming.

Finally, higher phonological overlap between the target and its translation equivalent predicted better accuracy, faster response times, and a higher likelihood of a language error than an omission[4]. This is consistent with the cognate facilitation effect observed in picture naming in both unimpaired bilinguals (e.g., Christoffels et al., 2006; Costa et al., 2005; Hoshino & Kroll, 2008) and bilinguals with aphasia (e.g., Lalor & Kirsner, 2001; Roberts & Deslauriers, 1999). This effect was equivalent in Spanish and in English, as seen by the absence of a significant interaction between phonological distance and language. Given that cognate effects in picture naming have been found to be stronger in the non-dominant or "weaker" language (e.g., Costa et al., 2000; Starreveld et al., 2014; Strijkers et al., 2010) but symmetrical in balanced bilinguals (e.g., Li & Gollan, 2021; Van Hell & Tanner, 2012), it seems that, in this respect, CA's performance approaches that of a more balanced bilingual.

Overall, CA's picture naming behaviours are generally consistent with what would be expected from an unimpaired late bilingual speaker. Namely, although language selection errors were the most frequent error type, factors that were predicted to increase cross-language activation (i.e., language mode and task (e.g., Grosjean, 1998), and phonological overlap between translation equivalents (e.g., Broersma, 2009; Broersma, Carter & Acheson, 2016; Broersma & De Bot, 2006)) were not related to an increase in this error type when compared to a correct response.

We conclude that there was no evidence that CA's cross-language errors were related to the need to inhibit the non-target language and disruption to this ability as a result of an impairment of inhibitory control (Abutalebi & Green, 2007; Green, 1998; Paradis, 2004). Indeed, in the Inhibitory Control model (e.g., Green, 1998), strongly activated representations in the non-target language require stronger inhibition as compared to representations in a weakly activated non-target language. In this framework, a disruption of inhibitory control mechanisms would have led to more interference from the non-target language, and this would be particularly prominent in conditions that maximise its level of activation. For CA, this would be when English was the target language, in bilingual language mode, in the subsequent-retrieval task rather than the picture naming task, and/or when translation equivalents had higher phonological similarity. This is not what we observed.

Rather than an inhibitory control impairment, for CA, we suggest that language selection errors most likely corresponded to lexical retrieval difficulties. This is consistent with research that emphasises the role of word-finding difficulties on occurrences of language mixing in bilingual aphasia (Goral et al., 2019; Lerman et al., 2019; Muñoz et al., 1999). In this context, language mixing errors are merely a specific type of paraphasia and may reflect a strategy (either conscious or unconscious) used to compensate for word-finding difficulties and maintain communication as efficiently as possible. Given that English and Spanish share a significant number of cognates, and that it is not exceptional for English speakers to know some Spanish or other Romance languages, uttering a Spanish translation equivalent when speaking English may indeed be a reasonable strategy for CA to improve communicative success.

In sum, the pattern of responses shown by CA encourages us to carefully consider the word “pathological” when describing the language mixing behaviours that are sometimes observed in bilingual aphasia. Indeed, Fyndanis and Lehtonen (2021) proposed that language mixing behaviours should only be deemed “pathological” if they happen in genuinely monolingual situations and result in communication breakdown. They also suggest that language mixing should only be viewed as a communicative strategy when the individual has knowledge that the communication partner is proficient in both the languages that are mixed (for related discussion, see Roberts, 2008, and Roberts and Deslauriers, 1999). While we agree with this in principle, we note that the situation is not always that clear-cut in the context of anomia. In such a situation where communication is already disrupted, language mixing occurrences, even when the communication partner does not speak the “other” language, may also have the potential to enhance communicativeness especially in cases when the two languages bear similarities, albeit also retaining the potential to disrupt communication. Whether communication will be disrupted or facilitated by the language mixing utterance is therefore not necessarily known in advance by the bilingual speaker with aphasia. In a situation where access to any word is compromised, the communicative value of language mixing behaviours should therefore be recognised rather than focusing solely on their disruptive potential.

Word count: 7,010 including the abstract.

*Acknowledgements
During the preparation of this article, Solène Hameau was funded by an Australian Research Council Discovery Program grant (DP190101490).

We thank CA for his enthusiastic participation in this study. We also thank Jaime Undurraga for valuable assistance with testing in Spanish, Macarena Moreno for help with stimulus preparation, assessment and response transcription, Josué Pino Castillo and Alejandra Romano for assistance in response transcription.

Disclosure statement: The authors report there are no competing interests to declare.

References

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, *20*(3), 242–275. https://doi.org/10.1016/j.jneuroling.2006.10.003

Ansaldo, A. I., Saidi, L. G., & Ruiz, A. (2010). Model-driven intervention in bilingual aphasia: Evidence from a case of pathological language mixing. *Aphasiology*, *24*(2), 309–324. https://doi.org/10.1080/02687030902958423

Bhat, S., & Chengappa, S. (2011). Language–mixing in bilingual aphasia: An Indian perspective. *South and Southeast Asian Psycholinguistics*, 372–380. https://doi.org/10.1017/CBO9781139084642.041

Bonfieni, M., Branigan, H. P., Pickering, M. J., & Sorace, A. (2019). Language experience modulates bilingual language control: The effect of proficiency, age of acquisition, and exposure on language switching. *Acta Psychologica, 193,* 160–170. https://doi.org/10.1016/j.actpsy.2018.11.004

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. In *Journal of the Royal Statistical Society. Series B (Methodological)* (Vol. 26, pp. 211–252). WileyRoyal Statistical Society. https://doi.org/10.2307/2984418

Brauer, M. (1998). Stroop interference in bilinguals: The role of similarity between the two languages. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 317–337). Psychology Press.

Broersma, M. (2009). Triggered codeswitching between cognate languages. *Bilingualism: Language and Cognition*, *12*(4), 447–462. https://doi.org/10.1017/S1366728909990204

Broersma, M., Carter, D., & Acheson, D. J. (2016). Cognate costs in bilingual speech production: Evidence from language switching. *Frontiers in Psychology*, *7*(SEP). https://doi.org/10.3389/fpsyg.2016.01461

Broersma, M., & De Bot, K. (2006). Triggered codeswitching: A corpus-based evaluation of the original triggering hypothesis and a new alternative. *Bilingualism: Language and Cognition*, *9*(1), 1–13. https://doi.org/10.1017/S1366728905002348

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service?. *Bilingualism: Language and Cognition, 13*(3), 359-371.

Chen, H., & Ho, C. (1986). Development of Stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(3), 397–401. https://doi.org/10.1037//0278-7393.12.3.397

Chen, H., & Leung, Y. (1989). Patterns of lexical processing in a nonnative language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(2), 316–325. https://doi.org/10.1037//0278-7393.15.2.316

Cheng, L., & Butler, K. (1989). Code-switching: A natural phenomenon vs language “deficiency”. *World Englishes*, *8*(3), 293–309. https://doi.org/10.1111/j.1467-971X.1989.tb00670.x

Christoffels, I. K., de Groot, A. M. B., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, *54*(3), 324–345. https://doi.org/10.1016/j.jml.2005.12.004

Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research*, *1147*(1), 192–208. https://doi.org/10.1016/j.brainres.2007.01.137

Colomé, À., & Miozzo, M. (2010). Which words are activated during bilingual word production? *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*(1), 96–109. https://doi.org/10.1037/a0017677

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning Memory and Cognition*, *26*(5), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, *41*(3), 365–397. https://doi.org/10.1006/JMLA.1999.2651

Costa, A., Roelstraete, B., & Hartsuiker, R. J. (2006). The lexical bias effect in bilingual speech production: Evidence for feedback between lexical and sublexical levels across languages. *Psychonomic Bulletin & Review*, *13*(6), 972–977. https://doi.org/10.3758/BF03213911

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language, 50*(4), 491–511. https://doi.org/10.1016/j.jml.2004.02.002

Costa, A., Santesteban, M., & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language*, *94*(1), 94–103. https://doi.org/10.1016/j.bandl.2004.12.002

Cuetos, F., Glez-Nosti, M., Barbón, A., & Brysbaert, M. (2011). SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicologica*, *32*(2), 133–143.

Dash, T., & Ansaldo, A. I. (2017). Clinical implications of neurocognitive control deficits in bilingual adults with aphasia. *Perspectives of the ASHA Special Interest Groups*, *2*(2), 117-125.

De Bot, K. (1992). A bilingual production model: Levelt's "speaking" model adapted. *Applied Linguistics*, *13*, 1–24.

Declerck, M., Koch, I., & Philipp, A. M. (2012). Digits vs. pictures: The influence of stimulus type on language switching. *Bilingualism*, *15*(4), 896–904. https://doi.org/10.1017/S1366728912000193

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801–838. https://doi.org/10.1037/0033-295X.104.4.801

Dewaele, J.-M. (2001). Activation or inhibition? The interaction of L1, L2 and L3 on the language mode continuum. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives. Vol. 31* (pp. 69–89). Multilingual Matters.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition,* 5(3), 175-197.

Downey, S. S., Sun, G., & Norquest, P. (2017). alineR: An R package for optimizing feature-weighted alignments and linguistic distances. *The R Journal, 9*, 138–152. https://doi.org/10.1080/09296170802326681

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six

European languages. *Quarterly Journal of Experimental Psychology (2006)*, *71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Evans, W. S., Hula, W. D., Quique, Y., & Starns, J. J. (2020). How much time do people with aphasia need to respond during picture naming? Estimating optimal response time cutoffs using a multinomial ex-gaussian approach. *Journal of Speech, Language and Hearing Research*, *63*(2), 599–614. https://doi.org/10.1177/00952443211055775

Fabbro, F. (2001). The bilingual brain: Cerebral representation of languages. *Brain and Language*, *79*(2), 211–222. https://doi.org/10.1006/brln.2001.2481

Fabbro, F., Skrap, M., & Aglioti, S. (2000). Pathological switching between languages after frontal lesions in a bilingual patient. *Journal of Neurology Neurosurgery and Psychiatry*, *68*(5), 650–652. https://doi.org/10.1136/jnnp.68.5.650

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression, 3rd Edition*. https://cran.r-project.org/web/packages/effects/effects.pdf

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language*, *43*(2), 182–216. https://doi.org/10.1006/JMLA.2000.2716

Fyndanis, V., & Lehtonen, M. (2021). Pathological language-switching/mixing and its relationship to domain-general cognitive control. In U. Røynneland, and R. Blackwood (Eds): *Multilingualism across the lifespan* (pp. 209-230). Routledge.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814. https://doi.org/10.1016/j.jml.2007.07.001

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory and Cognition*, *33*(7), 1220–1234. https://doi.org/10.3758/BF03193224

Goral, M., Levy, E. S., Obler, L. K., & Cohen, E. (2006). Cross-language lexical connections in the mental lexicon: Evidence from a case of trilingual aphasia. *Brain and Language*, *98*(2), 235–247. https://doi.org/10.1016/J.BANDL.2006.05.004

Goral, M., Norvik, M., & Jensen, B. U. (2019). Variation in language mixing in multilingual aphasia. *Clinical Linguistics & Phonetics*, *33*(10–11), 915–929. https://doi.org/10.1080/02699206.2019.1584646.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81. https://doi.org/10.1017/S1366728998000133

Green, D. W. (2011). Language control in different contexts: The behavioral ecology of bilingual speakers. *Frontiers in Psychology, 2*(103).

Green, D. W., & Abutalebi, J. (2008). Understanding the link between bilingual aphasia and language control. *Journal of Neurolinguistics, 21*(6), 558–576. https://doi.org/10.1016/j.jneuroling.2008.01.002

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515-530. https://doi.org/10.1080/20445911.2013.796377

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36*(1), 3–15. https://doi.org/10.1016/0093-934X(89)90048-5

Grosjean, F. (1997). The bilingual individual. *Interpreting. International Journal of Research and Practice in Interpreting*, *2*(1–2), 163–187. https://doi.org/10.1075/intp.2.1-2.07gro

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*(2), 131–149. https://doi.org/10.1017/S136672899800025X

Grosjean, F. (2001). The bilingual's language modes. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 1–22). Blackwell. https://doi.org/10.1002/9781405198431

Grosjean, F. (2008). Language mode. In *Studying bilinguals* (pp. 35–84). Oxford University Press.

Hameau, S., Biedermann, B., & Nickels, L. (2021). Lexical activation in late bilinguals: Effects of phonological neighbourhood on spoken word production. *Language, Cognition and Neuroscience*, *36*(4), 517–534. https://doi.org/10.1080/23273798.2020.1863438

Hameau, S., & Köpke, B. (2015). Cross-language transfer for cognates in aphasia therapy with multilingual patients: A case study. *Aphasie und Verwandte Gebiete (Internet)*, *2015*(3), 13–19. https://hal.archives-ouvertes.fr/hal-01248395/

Hermans, D., Ormel, E., van Besselaar, R., & van Hell, J. (2011). Lexical activation in bilinguals' speech production is dynamic: How language ambiguous words can affect cross-language activation. *Language and Cognitive Processes*, *26*(10), 1687–1709. https://doi.org/10.1080/01690965.2010.530411

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition*, *106*(1), 501–511. https://doi.org/10.1016/j.cognition.2007.02.001

Hoshino, N., & Thierry, G. (2011). Language selection in bilingual word production: Electrophysiological evidence for cross-language competition. *Brain Research*, *1371*, 100–109. https://doi.org/10.1016/j.brainres.2010.11.053

Ijalba, E., Obler, L. K., & Chengappa, S. (2012). Bilingual aphasia: Theoretical and clinical considerations. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism and multilingualism* (pp. 61–84). Wiley-Blackwell Publishing Ltd.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*(2), 277–288. https://doi.org/10.1016/j.actpsy.2007.06.003

Kohnert, K. (2004). Cognitive and cognate-based treatments for bilingual aphasia: A case study. *Brain and Language*, *91*(3), 294–302. https://doi.org/10.1016/j.bandl.2004.04.001

Kondrak, G. (2000). A new algorithm for the alignment of phonetic sequences. *Proceedings of the First Meeting of the North American Chapter of the Association for Computational Linguistics*, 288–295. http://dl.acm.org/citation.cfm?id=974343

Köpke, B., & Genevska-Hanke, D. (2018). First language attrition and dominance: Same same or different?. *Frontiers in Psychology, 9*(1963). https://doi.org/10.3389/fpsyg.2018.01963

Kroll, J. F., & De Groot, A. M. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In Li Wei (Ed.): *The bilingualism reader* (pp. 405-427). Taylor & Francis.

Kroll, J. F., & Dijkstra, T. (2010). The bilingual lexicon. In R.B. Kaplan (Ed.), *The Oxford handbook of applied linguistics, second edition* (pp. 349-370). Oxford University Press.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149–174. https://doi.org/10.1006/JMLA.1994.1008

Kroll, J.F., & Tokowicz, N. (2005). Models of bilingual representation and processing. In J.F. Kroll, & A.M.B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 531–553). Oxford University Press. http://www.citeulike.org/user/briordan/article/3081624

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition, 13*(3), 373-381. https://doi.org/10.1017/S136672891000009X

Kurland, J., & Falcon, M. (2011). Effects of cognate status and language of therapy during intensive semantic naming treatment in a case of severe nonfluent bilingual aphasia. *Clinical Linguistics and Phonetics*, *25*(6–7), 584–600. https://doi.org/10.3109/02699206.2011.565398

La Heij, W., Hooglander, A., Kerling, R., & Van der Velden, E. (1996). Nonverbal context effects in forward and backward word translation: Evidence for concept mediation. *Journal of Memory and Language, 35*(5), 648–665.

Lalor, E., & Kirsner, K. (2001). The role of cognates in bilingual aphasia: Implications for assessment and treatment. *Aphasiology*, *15*(10–11), 1047–1056. https://doi.org/10.1080/02687040143000384

Lenth, R. V. (2021). *emmeans : Estimated marginal means, aka least-square means.: Vol. R package* (1.6.3.). https://doi.org/10.1080/00031305.1980.10483031>.License

Lerman, A., Pazuelo, L., Kizner, L., Borodkin, K., & Goral, M. (2019). Language mixing patterns in a bilingual individual with non-fluent aphasia. *Aphasiology*, *33*(9), 1137–1153. https://doi.org/10.1080/02687038.2018.1546821

Li, C., & Gollan, T. H. (2021). What cognates reveal about default language selection in bilingual sentence production. *Journal of Memory and Language*, 118, 104214. https://doi.org/10.1016/j.jml.2020.104214

Mägiste, E. (1984). Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(2), 304–315. https://doi.org/10.1037/0278-7393.10.2.304

Mazzoni, D., Brubeck, M., & Haberman, J. (2005). *Audacity: Free audio editor and recorder*. http://audacity.sourceforge.net

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, *40*(1), 25–40. https://doi.org/10.1006/jmla.1998.2602

Mooijman, S., Schoonen, R., Roelofs, A., & Ruiter, M. B. (2021). Executive control in bilingual aphasia: A systematic review. *Bilingualism: Language and Cognition*, 1–16. https://doi.org/10.1017/s136672892100047x

Muñoz, M. L., Marquardt, T. P., & Copeland, G. (1999). A comparison of the codeswitching patterns of aphasic and neurologically normal bilingual speakers of English and Spanish. *Brain and Language*, *66*(2), 249–274. https://doi.org/10.1006/brln.1998.2021

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. John Benjamins.

Perecman, E. (1984). Spontaneous translation and language mixing in a polyglot aphasic. *Brain and Language*, *23*(1), 43–63. https://doi.org/10.1016/0093-934X(84)90005-1

R Core Team. (2014). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. http://www.r-project.org/.

Roberts, P.M. (2008). Issues in assessment and treatment for bilingual and culturally diverse patients. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related neurogenic communication disorders: Fifth edition* (pp. 245-276). Lippincott, Williams & Wilkins.

Roberts, P. M., & Deslauriers, L. (1999). Picture naming of cognate and non-cognate nouns in bilingual aphasia. *Journal of Communication Disorders, 32*(1), 1–23. https://doi.org/10.1016/S0021-9924(98)00026-4

Roberts, P. M., Garcia, L. J., Desrochers, A., & Hernandez, D. (2002). English performance of proficient bilingual adults on the Boston Naming Test. *Aphasiology*, *16*(4–6), 635–645. https://doi.org/10.1080/02687030244000220

Sadat, J., Martin, C. D., Magnuson, J. S., Alario, F.-X., & Costa, A. (2016). Breaking down the bilingual cost in speech production. *Cognitive Science*, *40*(8), 1911–1940. https://doi.org/10.1111/cogs.12315

Starreveld, P. A., De Groot, A. M. B., Rossmark, B. M. M., & Van Hell, J. G. (2014). Parallel language activation during word processing in bilinguals: Evidence from word production in sentence context. *Bilingualism*, *17*(2), 258–276. https://doi.org/10.1017/S1366728913000308

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: Electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*, *20*(4), 912–928. https://doi.org/10.1093/cercor/bhp153

Titone, D., Libben, M., Mercier, J., Whitford, V., & Pivneva, I. (2011). 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*, *37*(6), 1412–1431. https://doi.org/10.1037/a0024492

Treffers-Daller, J. (2011). Variability in code-switching styles: Turkish-German code-switching patterns. In R. Jacobson (Ed.), *Codeswitching Worldwide* (pp. 177–198). Mouton de Gruyter. https://doi.org/https://doi.org/10.1515/9783110812190.177

Tzelgov, J., Henik, A., & Leiser, D. (1990). Controlling stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(5), 760–771. https://doi.org/10.1037/0278-7393.16.5.760

Van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning*, *62*(SUPPL. 2), 148–171. https://doi.org/10.1111/j.1467-9922.2012.00710.x

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Venables, W.N., & Ripley, B.D. (2002). *Modern Applied Statistics with S. Fourth Edition*. Springer.

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, *20*(1), 6–10. https://doi.org/10.3758/BF03202594

Zheng, X., Roelofs, A., & Lemhöfer, K. (2018). Language selection errors in switching: Language priming or cognitive control? *Language, Cognition and Neuroscience*, *33*(2), 139–147. https://doi.org/10.1080/23273798.2017.1363401

Endnotes

[1] It is important to note that while these steps were to ensure strong activation of both languages, the resulting communication situation does not correspond to a bilingual mode in the sense Grosjean describes it, in that the participant was still explicitly required to respond in a particular language at a given time, and not the other language. A true bilingual mode would have meant that the participant was "allowed" to freely use either language, which was not the case here.

[2] The same analysis was performed on untransformed response times and showed the same pattern of significance, except for the interaction between language and phonological distance, which was now just under the significance threshold ($p = .075$).

[3] Analyses on untransformed response times led to the same pattern of significance.

[4] In an omission, no word form is sufficiently activated to reach a level that allows its selection. In that sense, a factor that makes it more likely to select a word in the other language than no word at all as is the case with an omission, can be seen as "facilitatory". Another source of evidence showing that a language error can be seen as facilitated retrieval as compared to an omission is the effect of frequency that was observed in CA: words were more likely to result in a language error than in an omission if they were more frequent (hence, "easier words" were more likely to result in a language error than in an omission).

Table 1. Summary of CA's language acquisition history and language use

	Spanish		English	
	Pattern of exposure	Proficiency	Pattern of exposure	Proficiency
Age 0: Birth in a South American country where Spanish is spoken				
Childhood/teenage years (age 0-18): Growing up in an exclusively Spanish-speaking environment.	Daily, from birth. At home, at school, in the community. Reading instruction in Spanish.	Typical of an educated monolingual speaker	No exposure (apart from hearing English songs on the radio)	No ability to use any English
Age 18 to 30 in an exclusively Spanish-speaking environment.	Daily At home, University, work, in the community, TV, newspapers, radio…	Typical of an educated monolingual speaker	English classes: - higher education: some English classes	Achieved very basic proficiency. Unable to have a conversation.
Age 30: Move to Australia with (Spanish-speaking) spouse				
Age 30 to 67: living, working and raising children in Australia.	Daily (~30% of the time) At home (wife, children), with some friends and close family. Some reading in Spanish (novels)	Maintained high proficiency	Upon arriving in Australia: 1 year of weekly English classes. Daily exposure (~70% of the time) At work, in the community, with friends (commonly mixed English and Spanish with some Spanish-	Achieved high proficiency, high confidence in English

			speaking friends), through TV / newspapers / books / internet… Some English at home with the family	
Age 67 (=retirement) to 72 (=stroke): retirement in Australia	Daily (75% of the time) At home (wife), with some friends and close family. Some TV/internet & reading novels in Spanish	Maintained high proficiency	Daily (25% of the time) In the community (grocery shopping, going to the doctor or to the bank), TV / newspapers / books / internet, with some friends, sometimes with family	Maintained high proficiency
Age 72 (=stroke) to 77 (=time of this study): living with aphasia in Australia.	Daily (75% of the time) At home (wife), with some friends and close family. Some TV/internet.	Moderate aphasia	Daily (25% of the time) Grocery shopping, going to the doctor or to the bank, aphasia activity group, some TV / internet, some friends, sometimes with family	Moderate aphasia

Table 2. Procedure: sequence of testing sessions by language

Session	Item set: 0-125	Item set: 126-250
Session 1 (monolingual)	English	
Session 2 (monolingual)		Spanish
Session 3 (monolingual)	Spanish	
Session 4 (monolingual)		English
Session 5 (bilingual, then subsequent-retrieval)	English, then Spanish	Spanish, then English
Session 6 (bilingual, then subsequent-retrieval)	Spanish, then English	English, then Spanish

Table 3. Accuracy and response latency across conditions

	Total (English + Spanish)			English			Spanish		
	N	% accurate	Mean RT(SD)	n	% accurate	Mean RT(SD)	n	% accurate	Mean RT(SD)
Monolingual	477	65	2035 (1330)	245	66	2224 (1442)	232	64	1827 (1166)
Bilingual	473	65	2305 (1551)	246	62	2255 (1569)	227	68	2349 (1539)
Subsequent-retrieval	444	61	2093 (1385)	226	61	1919 (1258)	218	61	2278 (1490)
All conditions	1394	64	2142 (1426)	717	63	2138 (1434)	677	64	2147 (1421)

n = total number of valid trials; RT(SD)= Mean and standard deviation of response latency in milliseconds

Table 4. Summary of the results of the logistic regression on accuracy, and multiple regressions on (transformed) response time.

	Accuracy				**RT: Monolingual and Bilingual**				**RT: Subsequent-retrieval**			
Predictors	*Odds Ratios*	*std. Error*	*Statistic*	*p*	*Estimates*	*std. Error*	*Statistic*	*p*	*Estimates*	*std. Error*	*Statistic*	*p*
(Intercept)	1.80	0.11	10.04	**< .001**	< 0.01	< 0.01	46.63	**< .001**	0.22	< 0.01	141.81	**< .001**
Log frequency	1.68	0.11	7.76	**< .001**								
Name agreement	1.16	0.07	2.44	**.015**								
Length	0.77	0.05	-3.73	**< .001**								
Visual Complexity	0.87	0.05	-2.30	**.022**								
Language	0.83	0.06	-2.70	**.007**	< 0.01	< 0.01	0.75	.451				
Phonological distance	0.81	0.05	-3.28	**.001**	< - 0.01	< 0.01	-2.11	**.035**	-0.01	< 0.01	-3.80	**< .001**
Condition					< 0.01	< 0.01	3.38	**.001**	N/A			
Condition*Language					< - 0.01	< 0.01	-3.02	**.003**	N/A			
Observations	1394				537				237			
R^2 Tjur	.086											
R^2 / R^2 adjusted					.062 / .053				.058 / .054			
AIC	1720.944				-5666.089				-1085.444			

Notes: Blank cells = predictors and interactions that were tested in the model but did not significantly improve the model's fit and where hence removed. N/A = refers to predictors/interactions that were never tested in this model.
Significant p-values (< .05) are in bold.

Table 5. Breakdown of response types, expressed as a) raw numbers, b) when applicable, as a percentage of errors (i.e., excluding acceptable alternatives and timing/sequence errors).

	Total		English		Spanish	
Response type name	Raw number	% errors (n=508)	Raw number	% errors (n=266)	Raw number	% errors (n=242)
correct	886	NA	451	NA	435	NA
acceptable alternative	72	NA	13	NA	59	NA
timing/sequence issue (subsequent-retrieval)	32	NA	19	NA	13	NA
correct in the non-target language	148	29	79	30	69	29
no response (omission)	126	25	59	22	67	28
Other errors*	234	45	128	48	106	44

* detailed in Appendix B

Table 6. Summary of the results of the logistic regressions on correct responses vs language errors, and on language errors vs omissions.

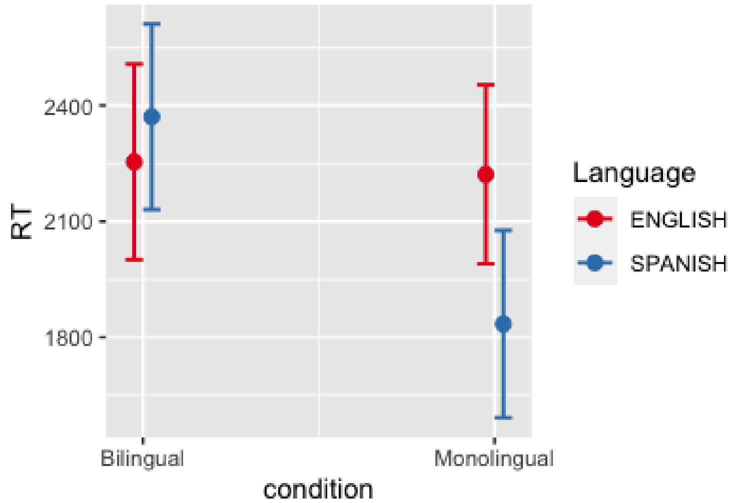
	Correct vs language error				**Language error vs Omission**			
Predictors	*Odds Ratios*	*std. Error*	*Statistic*	*p*	*Odds Ratios*	*std. Error*	*Statistic*	*p*
(Intercept)	7.25	0.80	17.88	**< .001**	1.60	0.28	2.69	**.007**
Log frequency	1.35	0.14	2.91	**.004**	1.56	0.29	2.35	**.019**
Length	0.76	0.08	-2.61	**.009**				
Language	0.81	0.08	-2.10	**.036**				
Phonological distance					0.50	0.09	-3.86	**< .001**
Condition (Bilingual)[1]	0.62	0.08	-3.49	**< .001**	2.63	0.59	4.28	**< .001**
Condition (Monolingual)[1]	0.65	0.09	-3.14	**.002**	3.42	0.82	5.12	**< .001**
Observations	1034				274			
R^2 Tjur	.046				.440			
AIC	809.194				250.939			

[1] Given that factors were sum-coded, the comparisons here are between a given level (here, Bilingual, or Monolingual) and the grand mean. The overall effect of condition as determined by a likelihood ratio test is: $X^2(2) = 29.893$, $p < .0001$ (correct vs language error), and $X^2(2) = 117.349$, $p < .0001$ (language error vs omission).

Note: Blank cells = predictors and interactions that were tested in the model but did not significantly improve the model's fit and were hence removed.

Significant p-values (< .05) are in bold.

Figure 1. Illustration of the significant interaction between condition and language on response time (monolingual and bilingual conditions only).



RT = Response latency (milliseconds)

Note: for better readability, this plot represents a model with untransformed response times.

Supplementary materials: one single file which includes:

Appendix A: Background testing results: this section contains a summary of the tests that were conducted with CA prior to the experimental part. It contains five parts: 1) A written summary of all the test results, 2) various individual tests, 3) CAT results, 4) BAT results, 5) Simon task results.

Appendix B: Response coding scheme: This section details the full coding scheme used to determine accuracy, response time and error types.

Appendix C: Model building steps: This section describes model-building steps from the initial full model to the final model reported in the article, for all reported analyses.

Supplementary materials

Hameau, S., Dmowski, U., & Nickels, L. (2022)

Factors affecting cross-language activation and language mixing in bilingual aphasia: A case study.

Appendix A (pp. 2-14)

Background testing results: this section contains a summary of the tests that were conducted with CA prior to the experimental part. It contains five parts: 1) A written summary of all background testing results, 2) various individual tests, 3) CAT results, 4) BAT results, 5) Simon task results.

Appendix B (pp. 15-22)

Response coding scheme: This section details the full coding scheme used to determine accuracy, response time and error types.

Appendix C (pp. 23-30)

Model building steps: This section describes model-building steps from the initial full model to the final model reported in the article, for all reported analyses.

Appendix A

A.1. Written summary of all background tests results

CA's language background was investigated through an informal interview and through the administration of the Language Experience and Proficiency Questionnaire (LEAP-Q: Marian et al., 2007). CA's first language was Spanish, and his second language was English which he had first learned formally through English classes as a student in higher education, and then at the age of 30, upon moving to Australia from a South American country where the official language is Spanish, with his (Spanish-speaking) spouse. CA reported that after the stroke, only English recovered for the first six months, and Spanish recovered thereafter. CA reported "equal" proficiency in English and Spanish (spoken and written) before the stroke. He reported having a "moderate" Spanish accent when speaking English. He rated his proficiency then as "perfect" in both languages in speaking, understanding spoken language and reading. CA's claim of "equal" and "perfect" proficiency in English is inconsistent with the likelihood that he would never have achieved native-like English grammar and pronunciation, and that he might have had broader vocabulary and use of structures in one language than the other one depending on the topic (e.g., greater work-related vocabulary in English than in Spanish). However, this claim suggests that CA was very confident speaking / reading / writing in his two languages in many different situations before his stroke. This can be explained by the fact that while he never stopped using Spanish, he was immersed in an English-speaking environment for over 50 years, performing a range of tasks requiring language in English, and using both languages daily. At the time of the study, CA reported being "dominant" in Spanish, and using more Spanish than English (75% / 25%). He spoke Spanish with most of his friends and family (sons, sister-in-law), and he used English in daily life when watching TV or browsing the Internet, going grocery shopping, to the bank, or to the doctor. This pattern of language use was the one AC adopted upon retiring 9 years prior to this study. In addition, after the stroke, AC would use English during a fortnightly activity group attended by peers with aphasia.

The term language dominance has been used to refer to different conceptions of the dominance construct (for a review, see Köpke & Genevska-Hanke, 2018). For instance, in adult bilinguals, dominance of one language often refers to increased proficiency in that language (e.g., Flege et al., 2002; Treffers-Daller, 2011; Wei, 2007). However, an alternative, more psycholinguistic definition of language dominance is based on the relative availability of each of the languages for processing (Gertken et al., 2014; Köpke & Genevska-Hanke, 2018). In this view, language dominance is interpreted in terms of processing facility (Birdsong, 2018) and is highly dependent on immediate language use context (Köpke & Genevska-Hanke, 2018). Therefore, using this definition of dominance, CA likely tended to be more Spanish-dominant at the time of the study.

To summarise, CA was a late, Spanish-English bilingual, with high pre-stroke proficiency in both languages and relatively Spanish-dominant.

Before the experimental procedure, background testing was administered, which included the Pyramids and Palm Trees Test (PPT; Howard & Patterson, 1992), the Comprehensive Aphasia Test (CAT; Swinburn et al., 2004), auditory synonym judgements (Kay, Lesser & Coltheart, 1996), the Boston Naming Test (BNT; Kaplan et al., 1987), and the Bilingual Aphasia Test (BAT, Paradis & Libben, 1987) in both Spanish and English. In addition, a Simon task was administered, targeting cognitive control abilities (for a review, see Lu & Proctor, 1995). Note that part of these assessments (i.e., the CAT and the BNT, and Part C of the BAT) were administered three years prior to this study, as part of another (unpublished) study. Although CA reported no change in his language or health status over the intervening period, it may be the case that his Spanish improved more than his English during that time, given that he was more exposed to Spanish than to English. This would not change the interpretation we make of the results of these assessments. English testing was administered by two native speakers of English (a speech pathology student, and for the earlier testing, an aphasia researcher), and Spanish testing was conducted, in separate sessions, by a native speaker of Chilean Spanish (trained speech pathologist/audiologist). Details of CA's performance on this set of background assessments is reported in Appendix A (Supplementary Materials).

CA performed within the normal range for all the cognitive (non-linguistic) tasks from the CAT's cognitive screen. In addition, his performance on the Simon task was compared to that of unimpaired bilingual and monolingual populations on the same task, taken from Bialystok et al. (2004) (see Table A.4 in Appendix A): CA performed within the expected range for older Spanish-English bilinguals, with relatively short reaction times overall and with a Simon effect in the lower range of his age category. He was faster, more accurate, and with a smaller Simon effect than Bialystok et al.'s (2004) healthy older monolinguals. This suggests good cognitive control abilities that are consistent with unimpaired bilingual performance. CA's conceptual processing abilities also seemed preserved as shown by his normal performance on the PPT. CA displayed relatively high scores in a range of subtests assessing the comprehension of single words, with a slightly better performance in Spanish than in English. However, syntactic comprehension was impaired in both English and Spanish, both with spoken and written sentences, as attested by CA's results on subtests of the CAT. CA also had trouble determining the grammatical correctness of sentences, in both of his languages (BAT grammaticality judgement subtest). CA's spoken output across various subtests was characterised by reduced fluency that was more apparent in English than in Spanish, some degree of anomia in both languages, and marked agrammatism in both languages. CA's agrammatism was also apparent in his reduced capacity to repeat and read aloud sentences in both languages, whereas repetition and reading aloud of single words were comparatively better preserved. CA often showed single-word intrusions of Spanish into English (and, to a lesser extent, of English to Spanish) in his spontaneous speech. Other single-word level errors (observed in picture naming subtests) included formal errors (e.g., “rump” for drum or “cama” for dama), phonologically related nonword errors especially on longer words (e.g., “sesokov” for stethoscope, or “diguitamente” for seguentemente) and sometimes conduite d’approche-type behaviours (“hang- han- hang- hanger”), semantic errors (“fence” for gate), mixed (semantic and phonological) errors (“harmony” for harp), and circumlocutions (“a stair down…electrical stair” for escalator).

Taken together, the single-word level background assessments suggested an impairment at the level of, or access to, the phonological output lexicon, with some impairment of the phonological output buffer. This impairment affected both of CA’s languages. Finally, based on CA’s language history, on his self-report of relative pre-morbid proficiency in both languages, and on the background test results, it seems that CA was experiencing parallel recovery of his languages at the time of the study.

References

Bialystok, E., Klein, R., Craik, F. I. M., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, *19*(2), 290–303. https://doi.org/10.1037/0882-7974.19.2.290

Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology, 9*(81). https://doi.org/10.3389/fpsyg.2018.00081

Flege, J. E., MacKay, I. R., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics, 23*(4), 567-598.

Gertken, L. M., Amengual, M., and Birdsong, D. (2014). Assessing language dominance with the bilingual language profile. In P. Leclercq, A. Edmonds, and H. Hilton (Eds), *Measuring L2 proficiency: Perspectives from SLA* (pp. 208-225). Multilingual Matters.

Howard, D., & Patterson, K. (1992). *The pyramids and palm trees test : A test of semantic access from words and pictures*. Pearson Assessment. https://latrobe.rl.talis.com/items/7FAA79F0-D149-2B81-FC62-94C7DE51BBBF.html

Kaplan, E., Goodglass, H., & Weintraub, S. (1987). *The Boston Naming Test (2nd ed.)*. Lea & Febiger. https://doi.org/10.1007/978-3-319-57111-9_869

Kay, J., Lesser, R., & Coltheart, M. (1996). Psycholinguistic assessments of language processing in aphasia (PALPA): An introduction. *Aphasiology*, *10*(2), 159–180. https://doi.org/10.1080/02687039608248403

Köpke, B., & Genevska-Hanke, D. (2018). First language attrition and dominance: Same same or different?. *Frontiers in psychology, 9*(1963). https://doi.org/10.3389/fpsyg.2018.01963

Lu, C.-H., & Proctor, R. W. (1995). of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review, 2*(2), 174–207.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research*, *50*(4), 940. https://doi.org/10.1044/1092-4388(2007/067)

Paradis, M., & Libben, G. (1987). *The assessment of bilingual aphasia*. Psychology Press. https://doi.org/10.4324/9781315802138

Swinburn, K., Porter, G., & Howard, D. (2004). *CAT : Comprehensive aphasia test*. Psychology Press. https://www.routledge.com/Comprehensive-Aphasia-Test/Swinburn-Porter-Howard/p/book/9781841693798

Treffers-Daller, J. (2011). Operationalizing and measuring language dominance. *International Journal of Bilingualism, 15*(2), 147-163.

A.2. Individual tests

Test	CA's score	Deviation from the norm
Pyramids and Palm Trees		
○ 3 pictures	52/52 (100%)	Within normal limits
○ 2 pictures + spoken word	48/52 (92%)	Within normal limits
○ 2 pictures + written word	51/52 (98%)	Within normal limits
PALPA 49 auditory synonym judgements	52/60 (87%)	No norms
Boston Naming Test (English Version)*	36/60	< 2 SD from the mean of adults aged 70-79
Average 1-min semantic word fluency (animals, vegetable, fruits) from the BAT		
Spanish	6	<1 SD below mean**
English	6	<1 SD below mean**
Average 1-min phonemic fluency (p,f,k)		
Spanish	3.6	<1 SD below mean**
English	1.3	<2 SD below mean**

* Test administered three years prior to the present study

** Norms from Rosselli et al. (2000) drawn from a population of older Spanish-English bilinguals from Florida.

Rosselli, M., Ardila, A., Araujo, K., Weekes, V. A., Caracciolo, V., Padilla, M., & Ostrosky-Solí, F. (2000). Verbal fluency and repetition skills in healthy older Spanish-English bilinguals. *Applied Neuropsychology*, *7*(1), 17-24.

A.3. Test battery in English: Comprehensive Aphasia Test (cognitive screen and language battery), administered three years prior to the present study.

Subtest	Score	Normal range
Cognitive screen		
Line bisection	0	-3.5 to 0.5
Semantic memory	10/10	9-10
Word fluency		
o animal	12	11-34
o /S/	4	0-25
Recognition memory	10/10	8-10
Gesture object use	11/12	8-12
Arithmetic	3/6	1-6
Language Battery		
Comprehension spoken words	15/15	13-15
Comprehension written words	13/15*	14-15
Comprehension spoken sentences	13/16*	14-16
Comprehension written sentences	11/16*	12-16
Comprehension spoken paragraphs	3/4	3-4
Repetition of words	12/16*	15-16
Repetition of complex words	3/3	3-3
Repetition of nonwords	4/5	2-5
Repetition of digit strings	4/7*	5-7
Repetition of sentences	4/6*	6-6
Naming objects	17/24*	21-24
Naming actions	2/5*	4-5
Spoken picture description	28*	33-87
Reading words	20/24*	22-24
Reading complex words	3/6*	6-6
Reading function words	6/6	4-6
Reading nonwords	4/10*	6-10
Writing: copying	27/27	24-27

Writing picture names	16/21	14-21
Writing to dictation	21/28*	22-28
Written picture description	18	18-66

*Performance outside of normal range

A.4. Bilingual Aphasia Test battery in both English and Spanish

Section	ENGLISH		SPANISH	
PART B: evaluation of language abilities in each language				
Spoken language processing / comprehension				
Lexical decision	27/30	90%	30/30	100%
Verbal commands: pointing to objects	10/10	100%	10/10	100%
Semantic acceptability	8/10	80%	10/10	100%
Verbal auditory discrimination (phonologically similar words)	15/18	83%	18/18	100%
Auditory word comprehension: odd-one out ("semantic categories")	5/5	100%	5/5	100%
Auditory word comprehension: synonym selection	5/5	100%	4/5	80%
Auditory word comprehension: antonym selection	8/10	80%	8/10	80%
Verbal commands: simple and semi-complex	10/10	100%	10/10	100%
Verbal commands: complex	9/15	60%	11/15	73%
Auditory syntactic comprehension	58/87	67%	72/87	83%
Grammaticality judgement	7/10	70%	4/10	40%
Spoken language production				
Automatic speech (series)	2/3	67%	3/3	100%
Verbal fluency: phonemic: p/f/k	2/2/0		4/3/4	
Verbal fluency: semantic: animals/vegetables/fruits	8/3/7		5/7/6	
Object naming	20/20	100%	19/20	95%
Generation of antonyms	8/10	80%	10/10	100%
Generation of morphologically derived words (adjectives from nouns)	6/10	60%	8/10	80%
Generation of morphologically derived opposites	7/10	70%	8/10	80%
Answering questions about an auditorily presented story	4/5	80%	4/5	80%

Sentence construction from imposed words	8/15		10/15	
Telling a story from a sequence of pictures	"A woman show the…bird sit on the tree branch almost no…carry on oh…yeah but no. Climb the tree eh nest. Fall down the tree by itself crushes the "accidente". Call the ambulance eh finish eh the…under the bed - no the bed- e sustain the broken arm. Feel sorry."		"La pal-paloma que estaba nel el árbol de que la señolo al jar- al al señalo al al el esposo que subió al árbol que que cepo al al rama, que la rama e que se espantó que la difícilmente que la… eh…la eh…tomó la tomó el nido que luego, que se caigó, que se caigó los pájaro i que se xxx que se … eh se… caigó al suelo y que secuentemente que la atención que la que lleva la a la l' « ampulancia » e que « diguidamente » la cama estaba con el pie « cobrado », que la sentida e hija que…la sentida faltando que sientide que el pájaro también sentida que los pajaritos lloraban. »	
Reading comprehension				
Reading comprehension for words	10/10	100%	10/10	100%

Reading comprehension for sentences	5/10	50%	4/10	40%
Written production				
Spontaneous writing : « write about your illness »	In *(month)* on *(date)* the sufer stoke. The hospital I took the 2-8 weeks. I dont is the aphasia is struch. The wife is the suffer my consentia		Se friete kstoke el año del *(year)*, el *(month)* atendio la ambacia. Que fue atende al hospital *(name of the hospital)* Que de alta y yo sabia que me afecto con la afasia.	
Transpositions				
Repetition of words and nonsense words	29/30	97%	27/30	90%
Repetition of sentences	2/7	29%	5/6	83%
Reading aloud: single words	6/10	60%	9/10	90%
Reading aloud: sentences	2/10	20%	3/10	30%
Copying	5/5	100%	5/5	100%
Writing to dictation: single words	3/5	60%	4/5	80%
Writing to dictation: sentences	21/27	78%	2/5	40%
Others				
Mental calculation (spoken prompt, and spoken output required)	5/15	33%	7/15	47%
PART C: translation abilities (administered three years prior to the present study)				
	Spanish->English		English->Spanish	
Pointing to word translation equivalents	5/5	100%	5/5	100%
Translating: words	10/10	100%	8/10	80%
Translating: sentences	4/18	22%	6/18	33%
	English		Spanish	
Grammaticality judgements of sentences	4/8	50%	6/8	75%
Correction of grammatically incorrect sentences	4/8	50%	5/8	62%

A.5. Simon task

	congruent		incongruent		
	Accuracy (%)	RT (in ms)	Accuracy (%)	RT (in ms)	Simon effect (in ms)
CA	100	579	95	664	85
Older bilinguals*	100	911 (374)	100	1659 (1151)	748 (807)
Younger bilinguals*	100	497 (252)	97	536 (273)	40 (32)
Older monolinguals*	99.2	1437 (561)	72.1	3150 (1310)	1713 (918)

* the older (monolingual and bilingual) and younger bilinguals scores were reported from Bialystok et al. (2004). Age range for the 20 older individuals (monolinguals and bilinguals) was 66 to 88 years old (mean= 71.9), and 30 to 54 years old (mean= 43) for the 10 younger bilinguals.

Note: Standard deviations are between parentheses.

Appendix B

Response coding scheme: This section details the full coding scheme used to determine accuracy, response time and error types.

Italics: target words

Italics, blue: example responses

Response considered: 1st complete attempt (so excluding interrupted syllables, hesitations/filler words like uh/eh). Exception for switch trials (see below).

Reminder of the 3 experimental conditions:

Monolingual condition example : Spanish trial

"raton"

Bilingual + subsequent retrieval condition example : Spanish-English trial

"raton , mouse"

Bilingual | Subsequent retrieval

ACCURACY CODING: 3 possibilities

CORRECT =1

When the intended response is correctly produced, in the correct target language.

- Includes a correct response produced with the addition of a modifier, provided that this is what the picture is actually showing.
  - *teacher* -> *female teacher*
  - *onion* -> *half the onion*
- Includes addition / deletion of a plural morpheme
  - *shoe* -> *shoes*
- Includes responses corresponding to the target but produced with an accent
  - *scale* -> */es'keɪl/*

- Includes correct responses produced after a hesitation consisting of the first phonemes of the target
  - *curtain* -> *curt-curtain*

Correct in bilingual / subsequent retrieval trials: the intended response has to be produced at the right moment in the sequence

  - *conejo, rabbit.* : in a Spanish-English trial with a picture of a rabbit

INCORRECT =0

Incorrect responses are further coded into error types (see below).

NA (trial omitted):

- Acceptable alternatives
  - *bike* -> *bicycle*
  - *aguacate* -> *palta* (= avocado)

- Timing/sequence:

  In 'bilingual + subsequent retrieval' trials when the correct response is produced but it is not clear which language was intended in that response (so it is unclear if this is a correct response, or a correct response in the non-target language).
  - Expected response: *cama-bed* -> CA's response: *bed no e cama*
    - ➔ no accuracy value for the target *bed* because although this response is given, it is here only produced when the intended response was in the other language, hence, it is impossible to know whether *bed* would have been produced accurately when actually trying to say that word in English, when intended. Note: in this case, the response for the target *cama* is considered incorrect (1st complete attempt, when *cama* is expected, is *bed* = correct in the non-target language).

Special case: when a response is repeated

The participant often repeats his response even if it is correct from the start:

This happens frequently in monolingual trials (which is not an issue for coding)

  - *bullet* -> bullet. bullet. bullet.

In a bilingual trial:

  - *coffee, café.* -> *coffee, coffee, café.*

- ➔ In such cases a rising intonation (coded with a comma here) after the 2nd *coffee* is taken as a sign that this is just a repetition of the word, and it is *café* that is considered the response to the 2nd target, NOT the 2nd instance of *coffee*. For this trial, we have: *coffee*: correct and with an RT, and *café*: correct but without an RT (because it is unclear whether the RT for that word should be calculated based on the 1st or the 2nd instance of *coffee*).

- ○ *tomato, tomate* -> *tomato, tomato. tomate.*
  - ➔ In this case, the second *tomato* has a falling intonation (coded with a full stop), so it is assumed that this is meant to be the response in Spanish (so it should be *tomate*). Even if subsequently the correct Spanish word is produced, this trial is incorrect (=correct word in the non-target language) so we have *tomato* = correct and with an RT, and <u>*tomate*</u> = incorrect (= correct response in the non-target language) so with no RT.

RESPONSE TIME CODING

RT: correct responses with a valid RT

<u>For either monolingual trials or bilingual trials:</u> time between the moment when the picture appears (signalled by a beep, so the onset of that beep) and the onset of the correct response. <u>For subsequent-retrieval trials:</u> time calculated from the onset of the previous response in the other target language (corresponding to the bilingual trial), and that of the (correct) response in the target language.

- ○ For *tiburón* in the 'bilingual + subsequent-retrieval' trial *shark, tiburón.* : RT for *tibúron* is calculated between the onset of *shark* and the onset of *tiburón.*

No RT:

- When correct = 0 or NA (see Accuracy coding)
- When correct =1 BUT RT is not reliable for either of the following reasons:
- word said after a hesitation
  - ○ *curtain* -> *curt-curtain*
- word preceded by a modifier
  - ○ *onion* -> *half the onion*

- word from previous trial overlapped with this trial (concurrently with the presentation of the picture).

- In subsequent retrieval trials: when the correct word is given but both when expected and when not expected:
    - *mouse* in a Spanish-English trial in which CA's response is *mouse, raton, mouse* (instead of *raton, mouse*)
        - ➔ Accuracy for mouse is 1 (based on the 2nd *mouse*, because it was correctly produced at the moment it was supposed to) but no RT is calculated (it is unclear how reliable the response time is because we do not know how much the production of the word was facilitated by its retrieval that just happened).

RESPONSE TYPE CODING

TL= target language

NTL= non-target language

Error type	Code TL	Code NTL	Description	Raw number of responses for each category”: Total (English; Spanish)
Correct	1	19	See above. Note: accuracy = 0 for a correct response in the non-target language.	886 (451; 435)
Acceptable alternative	10	Not observed	See above Accuracy = NA	72 (13; 59)
Phonological error	2	22	Error that is phonologically related to the target (the target and the response share at least 50% of phonemes, either 50% from the target or from the response). Phonemes that are accented count as the same as target phonemes (e.g., /a/ is accepted as equivalent to both /æ/ and /ɑː/): Additional coding: 2.1 Phonologically related nonword ○ *ostrich* -> */ˈɒster/* 2.2 Phonologically related existing word ○ *shield* -> *chill*	109 (69; 40)

Phonologically unrelated nonword error	3	n/a	Nonword response with less than 50% phonological overlap with the target: ○ *embarazada* (pregnant) -> */misise'sjɔn/*	2 (1; 1)
Semantic error	4	44	A noun that has a clear semantic relationship with the target: In the target language: ○ *arm* -> *leg* In the non-target language: ○ *crutches* -> *baston* (=cane)	39 (18; 21)
Mixed error	5	55	A semantic error that meets the criterion for phonological error too: In the target language ○ *factory* -> *farm* In the non-target language ○ *skull* -> *esqueleto*	11 (2; 9)
Semantic, then phonological error	6	66	A phonological error made on a semantic error: - In the target language: ○ *peineta* (=comb) -> */ses'pijo/* (phonological error on cepillo (=brush)) - In the non-target language:	4 (0; 4)

			○ *comb* -> */ses'pijo/*	
Circumlocution	7	Not observed	Multiword response that provides a characterisation of the target, or attempt to explain its function or purpose: ○ *crutches* -> *help the older people*	2 (1; 1)
Cognate error	8	Not observed	Only seen in English: when the participant takes an English word that resembles the Spanish translation equivalent of the target, as if it were a cognate, although it is not: ○ *balloon -> globe* (=globo in Spanish is the word for both balloon and globe). ○ *bow* -> *arch* (=arco in Spanish both means bow and arch)	4 (4; 0)
Morphological error	9	99	Morphologically related response (inflection or derivation) to the target: In the target language ○ *juggler* -> *juggle* In the non-target language ○ *swimmer* -> /*naða/* (nada) for nadador	14 (10; 4)

Visual error	11	1111	Participant names another visually-related item: ○ *skirt* -> *bag* OR response corresponds to a picture part (= participant names the wrong thing in the image): In the target language: ○ *mango* (=handle) -> *sartén* (=frying pan) In the non-target language: ○ *tatuaje* (=tattoo) -> *leg*	18 (7; 11)
Timing / sequence	12	n/a	In subsequent retrieval trials. The target response is given as the correct response but only at a time when a response in the other language is required. This means it is impossible to know whether the correct response would have been produced at the precise moment when it was required. ○ *bolsillo, pocket* -> *eh pocket, no.* Accuracy = NA	32 (19; 13)
Omission	13	n/a	No response except from filler words, automatisms, statements signalling that the participant doesn't know the response (e.g., *no sé*).	126 (59; 67)

Appendix C

Model building steps: This section describes model-building steps from the initial full model to the final model reported in the article, for all reported analyses.

C.1. Accuracy

Accuracy full model

glm(accuracy ~ Log frequency + Name agreement + Length + Visual Complexity + Language * Phonological distance * Condition, data = CA, family=binomial)

Model output:

	Accuracy			
Predictors	*Odds Ratios*	*std. Error*	*Statistic*	*p*
(Intercept)	1.81	0.11	10.02	**< .001**
Frequency	1.69	0.11	7.76	**< .001**
Name agreement	1.16	0.07	2.42	**.015**
Length	0.76	0.06	-3.75	**< .001**
Visual complexity	0.87	0.05	-2.29	**.022**
Language	1.45	0.20	2.72	**.007**
Phonological Distance	0.81	0.05	-3.29	**.001**
Condition: Monolingual vs Bilingual	1.13	0.19	0.73	.464
Condition: Bilingual vs Subsequent Retrieval	1.33	0.22	1.70	.089
Language * Phonological distance	1.02	0.12	0.14	.891
Language * Condition: Monolingual vs Bilingual	0.71	0.23	-1.04	.300
Language * Condition: Bilingual vs Subsequent Retrieval	1.16	0.38	0.44	.658
Phonological distance * Condition: Monolingual vs Bilingual	1.18	0.19	1.01	.314
Phononological distance * Condition: Bilingual vs Subsequent Retrieval	1.28	0.21	1.50	.133
Language * Phonological distance * Condition: Monolingual vs Bilingual	0.78	0.26	-0.75	.451

Language * Phonological distance * Condition: Bilingual vs Subsequent Retrieval	0.74	0.24	-0.91	.365
Observations	1394			
R^2 Tjur	0.092			
AIC	1730.476			

Accuracy model: analysis of deviance table

Computed with the command “Anova” from the package “car”. Type II tests were computed.

	LR Chisq	Df	p-value
Frequency	65.884	1	**< .0001**
Name agreement	5.855	1	**.015**
Length	14.162	1	**< .001**
Visual complexity	5.288	1	**.021**
Language	7.534	1	**.006**
Phonological Distance	10.880	1	**< .001**
Condition	2.856	2	.240
Language * Phonological distance	0.017	1	.897
Language * condition	2.271	2	.321
Phonological distance * condition	2.384	2	.303
Language * Phonological distance * condition	0.941	2	.625

The final resulting model (reported in the manuscript) includes the following predictors: Log frequency, Name agreement, Length, Visual complexity, and the experimental predictors Language and Phonological distance. The experimental predictor “Condition” was not retained in the final model as it did not contribute significantly to the model’s fit.

C.2. Response time

C.2.a. Response time: bilingual and monolingual conditions only

Response time: Bilingual and monolingual conditions only: full model

lm(RT ~ Log frequency + Name agreement + Length + Visual Complexity + Language * Phonological distance * Condition, data = CA.RT.Monolingual.Bilingual)

RTs are transformed: the BoxCox test pointed to raising RTs to the power of -0.75 as the appropriate transformation. This results in small (transformed) RT values and estimates in the model.

Model output

	(transformed) RT			
Predictors	*Estimates*	*std. Error*	*Statistic*	*p*
(Intercept)	72.77	1.06	68.85	**<.001**
Condition	5.03	2.05	2.45	**.014**
Language	-3.44	2.39	-1.44	.150
Phonological distance	2.96	1.08	2.73	**.007**
Name agreement	0.75	1.11	0.68	.498
Log frequency	-1.47	1.11	-1.33	.185
Visual complexity	1.93	1.01	1.91	.056
Length	0.33	1.39	0.24	.810
Condition * Language	9.66	4.10	2.35	**.019**
Condition * Phonological distance	1.71	1.97	0.87	.387
Language * Phonological distance	3.55	2.00	1.78	.076
Condition * Language * Phonological distance	-2.11	3.94	-0.53	.593
Observations	537			
R^2 / R^2 adjusted	0.053 / 0.034			
AIC	4933.225			

Response time: Bilingual and monolingual conditions only: analysis of deviance table

Computed with the command “Anova” from the package “car”. Type III tests were computed.

	Sum Sq	Df	F value	P value
(intercept)	1411311	1	2532.846	**< .0001**
Condition	3358	1	6.026	**.014**
Language	103	1	0.184	.668
Phonological distance	3777	1	6.779	**.009**
Name agreement	256	1	0.459	.498
Frequency	984	1	1.765	.184
Visual complexity	1038	1	3.657	.056

Length	32	1	0.058	.809
Condition * Language	3084	1	5.535	**.019**
Condition * Phonological distance	419	1	0.751	.386
Language * Phonological distance	442	1	0.794	.373
Condition * Language * Phonological distance	159	1	0.286	.593
Residuals	292536	525		

The final RT model (included in the manuscript) for the Monolingual and Bilingual conditions includes the 3 experimental variables, as well as a significant interaction between language and condition. None of the control variables significantly improved the model's fit.

C.2.b. Response time: Subsequent retrieval only

Response time: Subsequent retrieval only full model

RTs are transformed: the BoxCox test pointed to raising RTs to the power of -0.2 as the appropriate transformation. This results in small (transformed) RT values and estimates in the model.

lm(RT ~ Log frequency + Name agreement + Length + Visual Complexity + Language * Phonological distance * Condition, data = CA.RT.subsequent retrieval)

	RT			
Predictors	*Estimates*	*std. Error*	*Statistic*	*p*
(Intercept)	0.22	0.00	134.83	**< .001**
Language	-0.01	0.00	-1.48	.141
Phonological distance	-0.01	0.00	-3.72	**< .001**
Name agreement	0.00	0.00	0.28	.780
Log frequency	0.00	0.00	1.70	.090
Visual complexity	0.00	0.00	0.90	.370
Length	-0.00	0.00	-0.97	.332
Language * Phonological distance	-0.00	0.00	-0.89	.374
Observations	237			
R^2 / R^2 adjusted	0.107 / 0.080			
AIC	-1086.110			

Response time: Subsequent retrieval only analysis of deviance table

Computed with the command “Anova” from the package “car”. Type III tests were computed.

	Sum Sq	Df	F value	P value
(intercept)	10.441	1	18178.761	**< .0001**
Language	0.001	1	2.184	.141
Phonological distance	0.008	1	13.863	**< .001**
Name agreement	< 0.001	1	0.078	.780
Frequency	0.002	1	2.894	.090
Visual complexity	< 0.001	1	0.805	.730
Length	< 0.001	1	0.943	.332
Language * Phonological distance	< 0.001	1	0.793	.374
Residuals	0.1315	229		

Resulting RT model for the immediate translation task only has phonological distance as a predictor.

C.3. Language errors

C.3.a. Correct response vs language error

Correct response vs language error full model

glm(correct vs correct in the non-target language ~ Log frequency + Name agreement + Length + Visual Complexity + Language * Phonological distance * Condition, data = CA, family=binomial)

Model output summary

	Correct vs Correct in non-target language			
Predictors	*Odds Ratios*	*std. Error*	*Statistic*	*p*
(Intercept)	8.31	1.26	13.94	**< .001**
Frequency	1.34	0.14	2.86	**.004**
Name agreement	1.10	0.10	0.98	.327
Length	0.76	0.09	-2.46	**.014**
Visual complexity	0.85	0.08	-1.69	.091
Language	1.92	0.62	2.03	**.043**

Phonological Distance	1.15	0.15	1.10	.272
Condition: Monolingual vs Bilingual	0.10	0.05	-4.14	**< .001**
Condition: Bilingual vs Subsequent Retrieval	0.33	0.11	-3.25	**.001**
Language * Phonological distance	1.19	0.30	0.69	.487
Language * Condition: Monolingual vs Bilingual	0.30	0.33	-1.08	.280
Language * Condition: Bilingual vs Subsequent Retrieval	0.30	0.21	-1.74	.082
Phonological distance * Condition: Monolingual vs Bilingual	0.41	0.18	-2.01	**.044**
Phononological distance * Condition: Bilingual vs Subsequent Retrieval	0.69	0.20	-1.27	.206
Language * Phonological distance * Condition: Monolingual vs Bilingual	0.25	0.22	-1.55	.121
Language * Phonological distance * Condition: Bilingual vs Subsequent Retrieval	0.40	0.24	-1.56	.119
Observations	1034			
R^2 Tjur	0.057			
AIC	816.028			

Correct response vs language error analysis of deviance table

Computed with the command “Anova” from the package “car”. Type II tests were computed.

	LR Chisq	**Df**	**p-value**
Frequency	8.542	1	**.003**
Name agreement	0.942	1	.332
Length	6.002	1	**.014**
Visual complexity	2.890	1	.089
Language	4.891	1	**.027**
Phonological Distance	1.287	1	.257
Condition	34.230	2	**< .0001**
Language * Phonological distance	0.506	1	.477
Language * condition	3.610	2	.164
Phonological distance * condition	5.001	2	.082

Language * Phonological distance * condition	3.068	2	.216

The final resulting model (reported in the manuscript) for correct vs language errors includes the following predictors: Log frequency, Length, and the experimental predictors Language and condition.

C.3.b. Language error vs omission

Language error vs omission full model

glm(correct in the non-target language vs omission ~ Log frequency + Name agreement + Length + Visual Complexity + Language * Phonological distance * Condition, data = CA, family=binomial)

Model output summary

	Language error vs omission			
Predictors	*Odds Ratios*	*std. Error*	*Statistic*	*p*
(Intercept)	1.50	0.34	1.77	**.076**
Frequency	1.70	0.36	2.51	**.012**
Name agreement	0.94	0.16	-0.35	**.723**
Length	1.43	0.30	1.75	.081
Visual complexity	1.03	0.16	0.19	.852
Language	1.53	0.36	1.76	.078
Phonological Distance	0.44	0.10	-3.76	**< .001**
Condition: Monolingual vs Bilingual	2.85	0.80	3.75	**< .001**
Condition: Bilingual vs Subsequent Retrieval	4.60	1.48	4.73	**< .001**
Language * Phonological distance	0.97	0.21	-0.13	.900
Language * Condition: Monolingual vs Bilingual	0.62	0.17	-1.71	.087
Language * Condition: Bilingual vs Immediate Translation	1.19	0.38	0.55	.581
Phonological distance * Condition: Monolingual vs Bilingual	1.29	0.35	0.92	.357
Phononological distance * Condition: Bilingual vs Subsequent Retrieval	1.24	0.37	0.72	.474
Language * Phonological distance * Condition: Monolingual vs Bilingual	1.21	0.33	0.70	.487

Language * Phonological distance * Condition: Bilingual vs Subsequent Retrieval	0.54	0.16	-2.05	**.040**
Observations	274			
R^2 Tjur	0.472			
AIC	257.389			

Correct response vs language error analysis of deviance table

Computed with the command "Anova" from the package "car". Type II tests were computed.

	LR Chisq	Df	p-value
Frequency	6.709	1	**.010**
Name agreement	0.126	1	.723
Length	3.161	1	.075
Visual complexity	0.035	1	.852
Language	1.807	1	.179
Phonological Distance	12.957	1	**< .001**
Condition	118.884	2	**< .0001**
Language * Phonological distance	0.083	1	.773
Language * condition	1.204	2	.548
Phonological distance * condition	4.220	2	.121
Language * Phonological distance * condition	4.865	2	.088

The final resulting model (reported in the manuscript) for language errors vs omissions includes the control predictor frequency, and the experimental predictors Phonological distance and condition.