



Evidence for an integrated bilingual language system from discourse tasks in aphasia

Xuanyi Jessica Chen, Manuel J. Marte, Swathi Kiran & Esti Blanco-Elorrieta

To cite this article: Xuanyi Jessica Chen, Manuel J. Marte, Swathi Kiran & Esti Blanco-Elorrieta (02 Feb 2026): Evidence for an integrated bilingual language system from discourse tasks in aphasia, *Aphasiology*, DOI: [10.1080/02687038.2026.2624735](https://doi.org/10.1080/02687038.2026.2624735)

To link to this article: <https://doi.org/10.1080/02687038.2026.2624735>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 02 Feb 2026.



Submit your article to this journal [↗](#)



Article views: 611



View related articles [↗](#)



View Crossmark data [↗](#)

Evidence for an integrated bilingual language system from discourse tasks in aphasia

Xuanyi Jessica Chen^a, Manuel J. Marte^b, Swathi Kiran^b and Esti Blanco-Elorrieta ^{a,c}

^aDepartment of Psychology, New York University, New York, NY, USA; ^bCenter for Brain Recovery, Boston University, Boston, MA, USA; ^cDepartment of Neural Science, New York University, New York, NY, USA

ABSTRACT

Background: The extent to which bilingual individuals represent and process their two languages within a shared or partially distinct neural architecture remains a topic of ongoing debate. While both parallel and divergent patterns of impairment have been reported in bilingual aphasia, such findings likely reflect a spectrum of representational overlap influenced by dominance, proficiency, and task demands. Critically, few studies have examined how breakdown manifests across multiple levels of linguistic structure using ecologically valid, discourse-based tasks.

Aims: This study investigates whether Spanish-English bilinguals with aphasia exhibit parallel or dissociable patterns of impairment across their two languages, focusing on naturalistic narrative production and fine-grained analysis of speech error types and code-switching.

Methods & Procedures: Thirteen bilingual individuals with aphasia following acquired brain injury produced story retellings in both languages. Speech samples were transcribed and coded for phonological, morphological, syntactic, and semantic errors, and for the word type at which they occurred. Code-switches were also identified and categorized along the same dimensions. Analyses included generalized linear modeling, unsupervised clustering, and supervised machine learning methods.

Outcomes & Results: While participants made more errors in their non-dominant than in their dominant language, the structure and distribution of errors were highly similar across languages. Clustering algorithms and supervised classification analyses revealed that impairments were parallel across dominant and non-dominant languages. Code-switching occurred more frequently from the non-dominant to the dominant language, consistent with activation-based lexical selection.

Conclusions: Findings support an integrated bilingual language system that spans multiple levels of linguistic representation, modulated by language dominance. Naturalistic discourse tasks allow for richer characterization of bilingual language breakdown and may better inform both theoretical models and clinical management of bilingual aphasia.

ARTICLE HISTORY

Received 28 August 2025
Accepted 26 January 2026

KEYWORDS

Bilingualism; aphasia;
language organization

CONTACT Esti Blanco-Elorrieta  eb134@nyu.edu  Department of Psychology, 6 Washington Place, 588, New York University, 6 Washington Place, 588, New York, NY 10003, USA

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

1. Introduction

Understanding how bilingual individuals process and represent their two languages remains a central question in cognitive neuroscience, especially in the context of brain injury, where disruption to the language system offers a unique opportunity to infer its underlying architecture. At the heart of this inquiry lies a fundamental theoretical question: do bilinguals represent and process both languages within a single, unified neurocognitive architecture, or do they rely on partly distinct systems? A fully integrated model assumes that once a neural mechanism exists for a given linguistic operation (e.g., plural inflection), it serves that function across languages, treating cross-language variation similarly to within-language variation (e.g., across registers or dialects; Blanco-Elorrieta & Caramazza, 2021). In contrast, other perspectives argue that features such as formal grammatical structure (MacWhinney, 2022; MacWhinney et al., 2005), age of acquisition (Singleton, 2005; Singleton & Leśniewska, 2021), and usage context can give rise to at least partially distinct representations and processing routines. These views are not mutually exclusive, and growing evidence suggests that the bilingual mind exhibits both integration and separation depending on the linguistic domain and level of representation (Malik-Moraleda et al., 2024; Van de Putte et al., 2018).

The present study directly tests these accounts by examining how bilingual language breakdown unfolds across multiple linguistic levels following brain injury. Specifically, we investigate whether Spanish – English bilinguals with post-stroke aphasia show parallel or dissociable patterns of impairment across their two languages, and whether code-switching behavior provides converging evidence for an integrated or partially distinct language system. By analyzing naturalistic, discourse-based production, we aim to capture how bilingual language organization manifests in people with aphasia under real-world communicative conditions rather than decontextualized laboratory tasks.

Within healthy populations, much prior work has focused on the lexico-semantic level (Chee et al., 1999; Dehaene et al., 1997; Hasegawa et al., 2002; Molinaro et al., 2024; Yang et al., 2017), with relatively limited attention to morphosyntactic and phonological representations (Declerck, Kleinman, et al., 2020; Ou et al., 2020). While conceptual representations are generally believed to be shared across languages (Chen et al., 2024; Van de Putte et al., 2018), debates continue over whether morphology, syntax, and phonology are supported by common or distinct neural mechanisms. Findings vary from overlapping activation patterns for L1 and L2 (Musso et al., 2003; Perani & Abutalebi, 2005; Sakai et al., 2004; Tettamanti et al., 2002; Yang et al., 2017) to evidence of differential activity for L1 and L2 in distinct brain regions (Golestani et al., 2006; Ip et al., 2017; Lehtonen et al., 2009), and even language-specific differentiation across regions of interest (Gao et al., 2023; Meykadeh et al., 2021; see; Cargnelutti et al., 2019 for meta-analysis; Biondo et al., 2023 for review).

Studying bilinguals with post-stroke aphasia provides a rare opportunity to test this directly: lesions to shared neural resources should lead to parallel impairments in both of a bilingual's languages, whereas lesions affecting language-specific regions should yield selective deficits in one language only. In this way, bilingual aphasia serves as a natural lesion model that can arbitrate between integrated and partially distinct accounts of bilingual language representation. By analyzing the nature and distribution of errors

across linguistic levels, we can infer whether both a bilingual's languages rely on common processing mechanisms or on separable neural subsystems.

In addition to linguistic level, word class provides a finer-grained window into the organization of bilingual language representations. A large body of prior neuropsychological and neuroimaging work has shown that aphasia can differentially affect content words (nouns, verbs, adjectives) and function words (articles, prepositions, pronouns, conjunctions), reflecting their distinct roles in language production and neural representation (Bastiaanse & Thompson, 2012; Caramazza & Zurif, 1976; Friederici, 2011). Content words carry rich semantic and conceptual information, whereas function words primarily encode grammatical relations and structural dependencies (Friederici, 2011). Although these distinctions have often been discussed in terms of partially separable neural substrates in prior work (e.g., Hagoort, 2005; Tyler et al., 2011), the present study does not aim to make claims about neural localization. Rather, we adopt this distinction as a theoretically motivated representational dimension for behavioral analysis. In bilingual speakers, this distinction may carry additional relevance: content words may plausibly share cross-linguistic representational nodes due to their grounding in conceptual – semantic knowledge, whereas function words encode language-specific grammatical distinctions (e.g., agreement, case, gender) that may differ across languages or be absent in one language altogether. This motivates examining word class as a key dimension along which bilingual language breakdown may show both convergence and asymmetry.

Clinically, this distinction is also evident: agrammatic and non-fluent aphasias often involve disproportionate difficulty with function words and inflectional morphology, while fluent aphasias tend to spare grammatical morphemes but produce semantically empty content (Bastiaanse & Thompson, 2012; Caramazza & Zurif, 1976). Because bilingual aphasia might manifest either as parallel or dissociable breakdown across these lexical categories, analyzing errors by word class offers a more precise diagnostic lens on the structure of the bilingual language system. If a bilingual's two languages share a unified representational and control architecture, we would expect similar function – content asymmetries in both the dominant and non-dominant language; if they are partly distinct, the relative vulnerability of these categories could diverge. Incorporating word class thus allows us to examine not only whether the two languages break down in parallel, but how deeply that parallelism extends into the grammatical versus lexical strata of language.

Foundational neuropsychological work on bilingual aphasia has documented both parallel and differential patterns of post-stroke impairment (Green, 2005), and these contrasting outcomes have been central to the debate over whether bilingual language is supported by shared or partially distinct systems. Classic reports by Paradis (2001) and by Fabbro (2001) described cases in which one language was disproportionately or selectively impaired, in some instances with marked dissociations at specific linguistic levels (e.g., lexical access impaired in one language but preserved in the other, or grammatical morphology disrupted in only one language), and in some cases with asymmetric recovery trajectories across languages over time. Such cases have been interpreted as evidence that, at least under some lesion and experience profiles, the two languages can rely on partly independent neural resources and/or processing routines. In contrast, group-level work has mainly found broadly parallel impairment across languages in bilingual persons with aphasia, such that both languages show qualitatively

similar error types and breakdown patterns, even when they differ in overall severity (e.g., more errors in the less proficient or less frequently used language; Goral et al., 2006; Green et al., 2010; Kiran & Iakupova, 2011; Kohnert, 2004). On this view, apparent between-language differences largely reflect graded accessibility (frequency, dominance, pre-morbid use) rather than distinct representational architectures. Together, these literatures suggest that bilingual aphasia outcomes span a continuum from tightly coupled to partially dissociable, and that this continuum is influenced by factors such as age of acquisition, dominance, lesion site, and post-stroke language use (Kiran et al., 2013; Peñalosa & Kiran, 2019; Peñalosa et al., 2019).

Despite the richness of this work, two key gaps remain and directly motivate the present study. First, most prior studies have focused on a narrow set of tasks (e.g., picture naming, translation, word-level repetition) and therefore speak primarily to lexico-semantic retrieval. Much less is known about whether bilinguals show parallel versus dissociable impairment profiles across multiple levels of linguistic representation (i.e., phonology, morphology, syntax, and semantics) when producing connected speech.

Second, while a substantial body of work has examined code-switching and language mixing in bilingual aphasia, particularly in spontaneous discourse (e.g., Goral et al., 2006; Green et al., 2010; Lerman et al., 2019, 2020, 2025; see also references therein) this literature has primarily focused on the communicative functions, psycholinguistic triggers, and clinical profiles associated with language mixing. For example, Goral and colleagues have demonstrated that code-switching in aphasia reflects strategic responses to lexical retrieval difficulty and varies as a function of premorbid dominance, recovery trajectory, and task demands (Goral et al., 2006; Lerman et al., 2019, 2020, 2025). Collectively, this research has established that code-switching in aphasia is systematic, meaningful, and sensitive to both impairment and recovery.

What remains less explored, however, is whether the structural properties of spontaneous code-switching can be leveraged as a diagnostic behavior to adjudicate between competing models of bilingual language organization. That is, beyond characterizing when and why bilingual individuals with aphasia switch languages, relatively few studies have asked whether the direction, lexical locus (i.e., word class), and representational level of switches provide principled evidence for integrated versus partially distinct bilingual language systems.

This question is theoretically consequential because code-switching is often described clinically as a symptom of impaired control, yet it may instead reflect adaptive recruitment of whichever lexical or structural resource is most available at a given moment. As such, this behavior provides a dynamic window into the functional organization of the bilingual system. Because code-switches occur at specific lexical items or phrases, examining their distribution across word classes (e.g., content vs. function words) provides an additional lens on whether bilingual lexical selection and control operate similarly across grammatical categories. If a bilingual's two languages draw on shared neural substrates, switches should emerge as fluid transitions within a single integrated network rather than as costly shifts between distinct systems. Under activation-based accounts, this predicts that switches will preferentially occur from the non-dominant to the dominant language (i.e., the language with overall greater activation; Blanco-Elorrieta & Caramazza, 2021). This prediction is consistent with behavioral evidence from neurotypical

bilinguals, where spontaneous and voluntary code-switching in naturalistic contexts tends to favor switches into the dominant or more accessible language, particularly under increased lexical difficulty or reduced planning time (e.g., Blanco-Elorrieta & Pylkkänen, 2017; Gollan & Ferreira, 2009; Kleinman & Gollan, 2018). Such findings support the view that switching reflects graded availability within a shared system rather than failures of control.

In contrast, partially separated or inhibitory models (Green, 1998) propose that producing the non-dominant language requires active suppression of the dominant one, such that failures of inhibition result in “reverse dominance”, i.e., more frequent switching into the non-dominant language (Declerck, Kleinman, et al., 2020; Fadlon et al., 2019; Goldrick & Gollan, 2023; Gollan et al., 2014; Li & Gollan, 2018). Behavioral evidence for this account has primarily emerged from paradigms that explicitly tax language control, such as forced-switch or mixed-language tasks, where increased switching into the non-dominant language has been observed (Fadlon et al., 2019; Gollan et al., 2014; Li & Gollan, 2018). Examining the direction and frequency of code-switches in bilingual PWA thus provides a behavioral proxy for how tightly coupled or segregated their two languages are at the neural level. The current study addresses these gaps by analyzing naturalistic narrative production in Spanish – English bilinguals with aphasia, coding error types across multiple representational levels and examining the directionality and function of spontaneous code-switches. By integrating error analysis and code-switching within a single discourse-based framework, we test whether bilingual language breakdown reflects an integrated or partially distinct representational architecture.

Language typology may also play a role in shaping bilingual language organization. Studies comparing typologically distant language pairs (e.g., English – Mandarin, Spanish – Arabic) suggest that linguistic distance can influence cross-language transfer, representational overlap, and recovery profiles (Ansaldo & Saidi, 2014; Ghazi-Saidi & Ansaldo, 2017). That said, if the principles of language representation are indeed universal, as posited by many neurocognitive models, then typological distance alone should not necessarily result in fundamentally different neural architectures. At present, no clearly articulated theory specifies how or why structural differences in morphology, syntax, or phonology would causally give rise to distinct neural or cognitive organizations. This remains an important area for future theoretical development. The current study focuses on Spanish and English, two Indo-European languages that, while different in grammatical and phonological properties, share structural similarities that may facilitate integration (e.g., subject-verb-object word order, alphabetic script).

Crucially, regardless of typological distance, different theoretical models make distinct predictions about how linguistic deficits and code-switching behaviors (i.e., alternating between languages within a conversation or utterance) should appear in bilingual PWA. Integrated models (e.g., Blanco-Elorrieta & Caramazza, 2021; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982) predict that an impairment at a given representational level should affect both languages similarly in quality, with differences arising only in degree (for example, greater impairment in the non-dominant language due to less frequent use; Nadeau, 2019, or lower proficiency; Peñaloza, Dekhtyar, et al., 2020). Under this view, parallel error profiles would be expected across languages, even if severity varies (Jacquemot et al., 2007; J; Jefferies & Lambon Ralph, 2006; Lallini et al., 2007; Martin et al., 1996). Crucially, these predictions must be evaluated separately for each linguistic

level: it is entirely plausible that phonological representations might be integrated while morphosyntactic representations remain distinct.

In contrast, partially separated models predict that impairments can show language-specific dissociations – such as impaired phonology in one language but intact phonology in the other, or distinct grammatical error profiles (MacWhinney, 2022; Paradis, 2004). Crucially, such dissociations have been observed in single-case reports (e.g., Fabbro, 2001; Paradis, 2001). Thus, our aim here is to test the extent to which integration holds across linguistic levels using naturalistic production tasks.

In all, the hypotheses we test are as follows.

- (1) If representations are fully integrated, error patterns should be qualitatively identical across a bilingual's dominant and non-dominant languages, differing only in quantity based on dominance or proficiency. This parallelism should hold both across linguistic levels (phonology, morphology, syntax, lexico-semantics) and across word classes, such that any content – function asymmetries (or noun/verb differences) are expressed similarly in both the dominant and non-dominant language, regardless of whether that language is Spanish or English.
- (2) If representations are partially separated, we should observe Language \times Linguistic Level interactions, with some levels showing parallel deficits and others showing dissociations, and/or divergence by word class (e.g., different relative vulnerability of content vs. function words, or different noun/verb profiles across languages), with some domains showing parallel deficits and others showing dissociations.
- (3) If activation-based lexical selection governs bilingual production, code-switching should predominantly occur from the non-dominant to dominant language. Alternatively, under inhibitory accounts, producing the non-dominant language requires active suppression of the dominant language; thus failures of inhibition should yield “reverse dominance”, with code-switching occurring more frequently from the dominant to the non-dominant language.

In the current study, we aim to characterize bilingual aphasia by analyzing the typology of language impairments and the direction and function of code-switches during narrative production. Rather than relying on decontextualized tasks (e.g., picture naming), we use open-ended storytelling to elicit discourse that spans multiple representational levels and allows us to examine real-time language use across both languages. This approach provides an ecologically valid and theoretically rich opportunity to trace patterns of impairment and code-switching across specific linguistic domains. By doing so, we aim to highlight possible areas of integration and dissociation, and consider how these patterns might inform models of bilingual language organization.

2. Methods

2.1. Participants and experimental procedure

We collected data from 13 English-Spanish bilinguals with aphasia (9 F, 4 M; age 26–68; $M = 49$, $SD = 13.74$). All PWA were living in the United States at the time of the study. Their language profile was assessed with the Language Use Questionnaire (LUQ; Kastenbaum et al., 2019;

Marte et al., 2022), a notably comprehensive test that measures daily language use, language proficiency of family members, language of instruction at each level of education, language exposure, language confidence and self-reported language ability in hearing, speaking, reading and writing across their lifetime (see Figure 1(a) top panel).

All PWAs were tested during the subacute and chronic phase of their stroke and completed the Aphasia Bank protocol (MacWhinney et al., 2011), which aims to elicit PWA's speech by asking them to: i) describe their stroke experience and a memorable event, ii) describe some pictures (the Broken Window sequence, the Refused Umbrella, the Cat in the Tree), iii) tell the Cinderella story, and iv) describe the steps required to make a Peanut Butter and Jelly sandwich. All testing was administered by trained, bilingual Spanish-English speech-language pathologists and research assistants. Each testing session was conducted entirely in one language (either English or Spanish), with examiners using only the target language during task administration to maintain consistent

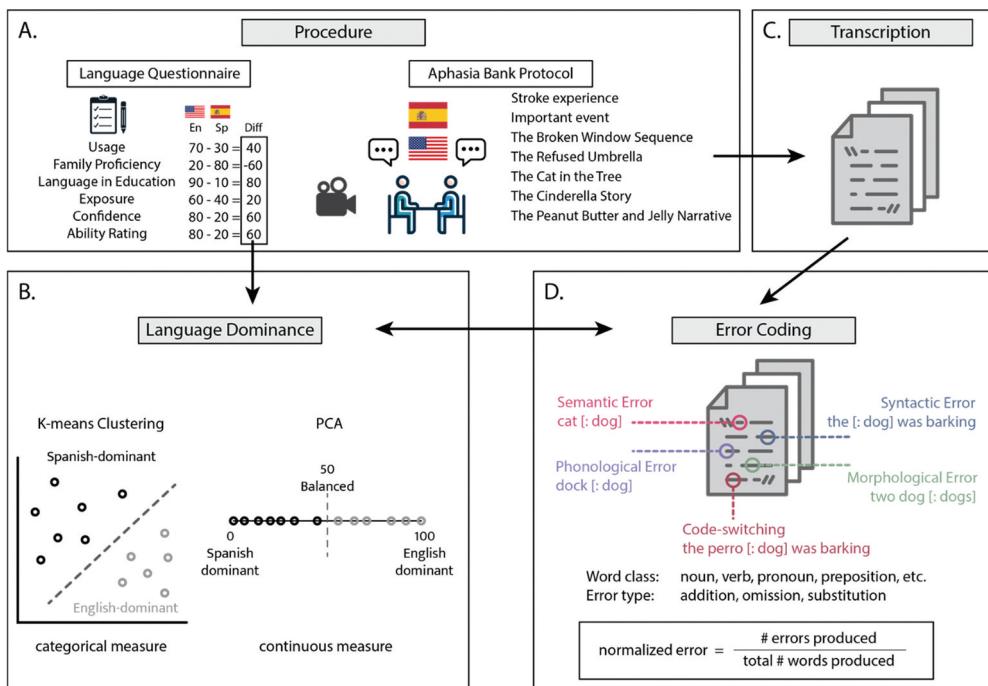


Figure 1. (A) testing and analyses procedure. During the testing session, persons with aphasia (PWA) completed the language use questionnaire (LUQ; Kastenbaum et al., 2019) and the aphasia bank protocol (MacWhinney et al., 2011) in English and Spanish. (B) Language dominance calculation. The difference between languages in every measure of the language questionnaire was calculated and used to derive two measures for language dominance: a categorical measure of language dominance using K-means clustering and a continuous measure using PCA. (C) Transcription. PWAs' speech was transcribed by two bilingual native speakers. (D) Error coding. PWAs' speech was coded for errors and code-switches. For each error, we coded for the level of representation (semantic, syntactic, phonological, morphological), word class (noun, verb, adjective, adverb, pronoun, determiner, preposition, conjunction), and error type (addition, omission, substitution). Normalized errors were calculated by scaling the raw frequency of errors by the total number of words produced (excluding repetition and retraction).

Table 1. Demographics and clinical characteristics of the PWA ($n = 13$). Values include time post-stroke (months) and western aphasia Battery-revised aphasia quotient (WAB AQ) scores in both English and Spanish, with the difference between languages (English minus Spanish).

	Age	Time post-stroke (months)	WAB_AQ English	WAB_AQ Spanish	WAB_AQ English-Spanish
Mean	49.5	57.08	76.85	79.38	-2.54
Standard Deviation	13.18	62.78	12.22	10.10	16.63
Range	27.4 – 68.3	7 – 244.7	61.7 – 97.2	60.5 – 92.6	-23.7 – 29.1

language context. Testing languages were counterbalanced across sessions, and examiners were instructed to minimize verbal feedback during narrative production to avoid influencing participants' spontaneous language production and code-switching behaviors. The speech output of these tasks constitutes the corpus for our analyses. All participants produced speech samples larger than 300 words in each language, beyond the minimum established to provide test-retest stability of connected speech measures for adults with aphasia (Brookshire & Nicholas, 1994). Ten patients had confirmed left-hemisphere lesions, while three had unspecified lesion lateralization in their medical records. Etiologies included ischemic stroke ($n = 5$), hemorrhagic stroke ($n = 1$), unspecified CVA ($n = 5$), Moyamoya disease ($n = 1$), and brain tumor ($n = 1$). The Western Aphasia Battery (WAB; Kertesz, 2007) was additionally completed to determine the severity of the aphasia in both languages. All PWAs received a WAB aphasia quotient score of ≥ 60 in both languages, indicating that they were mild or moderate aphasia per WAB classification (see Table 1). WAB testing was counterbalanced across languages and sessions to control for order effects. All tests were administered by a trained speech-language pathologist either in person or on zoom.

2.2. Language dominance calculation

Critical to our hypothesis testing is to determine i) whether there is a *quantitative* difference across languages that reflects higher resilience to damage of higher frequency elements (here understood also as dominant language elements), and ii) whether a *qualitatively* different profile of errors emerges between dominant and non-dominant language. Both of these require defining each PWA's pre-morbid language dominance, which is non-trivial due to the varying nature of every person's language experience.

Traditionally, Age of Acquisition and/or proficiency have been used to operationalize language dominance, but they are too one-dimensional to capture the multidimensional nature of language experience. To address this, we adopted a comprehensive, data-driven approach by incorporating all dimensions from the Language Use Questionnaire (LUQ; Kastenbaum et al., 2019; Marte et al., 2022), including language usage, exposure, proficiency of family members, educational language history, confidence, and self-rated ability across modalities. We calculated the difference between English and Spanish for each dimension, and then applied both K-means clustering and Principal Component Analysis (PCA) to these difference scores to classify participants' relative language dominance (Figure 1(b)).

K-means clustering is an algorithm that takes multidimensional data (in our case, the data from every language experience dimension for each subject) and groups it on

K number of clusters based on similarity across all dimensions. Here, we used the algorithm to divide the data into two clusters: Spanish-dominant and English-dominant. The decision to apply a two-cluster ($K = 2$) K-means solution was theoretically motivated. Our core predictions center on asymmetries between a speaker's dominant and non-dominant language, predictions that hold even when the dominance is subtle. Unless a participant is exactly balanced, even small asymmetries in language experience are theoretically meaningful and may shape processing outcomes. In fact, including participants with marginal dominance differences provides a more stringent test of our hypotheses. If effects emerge across this full range, they are less likely to be artifacts of selecting only clearly unbalanced individuals and more likely to reflect generalizable patterns of bilingual processing.

K-means clustering grouped participants into two data-driven categories. The algorithm classified 7 participants in the Spanish-dominant group and 6 in the English-dominant group, based on their composite language profiles.

To preserve the continuity of dominance and to validate the clustering outcome, we also applied PCA, which reduces multidimensional data into a continuous scale. PCA is a linear dimensionality reduction technique, which takes high-dimensional data (i.e., the data from every dimension in the language experience questionnaire) and transforms it into low-dimensional data by minimizing redundant information across related measures. In our case, we want to obtain one variable from the language experience questionnaire that reflects subjects' relative language dominance. We first calculated the difference between English and Spanish in each of the language experience dimensions (e.g., proficiency), and submitted the vector containing these difference scores to a PCA analysis. Then, we took the first component, which is the component that best explains the variance in the data (in our case, 88.86%), and we obtained a raw language dominance measure by calculating the projection of all dimensions for each subject along this first component. To determine the directionality of this measure (i.e., whether positive indicates English-dominant or Spanish-dominant), we inspected the language profile of the two subjects who received the maximum and minimum values. Finally, we scaled and recentered the raw measures to the range of 0–100, such that the most Spanish-dominant measure is mapped to 0 and the most English-dominant measure is mapped to 100. The value represents the relative dominance of English or Spanish, based not just on Age of Acquisition or Proficiency, but rather based on the compound of all the answers they provided in the comprehensive Language Use Questionnaire (LUQ; Kastenbaum et al., 2019).

Importantly, the results across the categorical and continuous measures in our analysis were fully consistent: participants with a score of 50 or more in PCA were clustered in the English-dominant cluster by K-means, and those with a score of less than 50 clustered as Spanish-dominant. This strong convergence across methods enhances confidence in our classification scheme, while also preserving the ability to explore graded dominance effects. This dual approach ensures both theoretical precision and empirical robustness in our treatment of language dominance.

2.3. Data transcription and coding

Recordings of the testing sessions were transcribed in the Computerized Language Analysis Software (CLAN; MacWhinney & Wagner, 2010; MacWhinney, 2017; see

Figure 1 for a detailed description of the procedure). All recordings were independently transcribed and coded following the AphasiaBank transcription guidelines by native speakers of each language. To assess inter-rater reliability, we followed Koo and Li's (2016) guidelines for selecting and reporting intraclass correlation coefficients (ICC). Specifically, we calculated a two-way random-effects, absolute-agreement, single-measure ICC, which is appropriate when each subject (here, each transcript) is independently rated by two coders randomly selected from a larger pool. ICC values were computed for both transcription accuracy and error categorization across all linguistic levels (phonological, morphological, syntactic, and lexico-semantic). According to Koo and Li (2016), ICC values below .50 indicate poor reliability, between .50–.75 moderate, between .75–.90 good, and above .90 excellent reliability. Prior to formal analysis, coders independently transcribed and coded 20% of the dataset (randomly selected across participants and languages). ICCs were then computed for (a) number of words transcribed, (b) number of coded errors per linguistic level, and (c) classification consistency of error types. All reliability estimates exceeded the threshold for *good* reliability (ICC > .80).

To the extent that AphasiaBank was created for English, the range of error codes did not cover all the possible errors one could encounter in Spanish. Thus, we developed and expanded this guide to incorporate all the types of mistakes PWAs could make in Spanish. Importantly, language-specific surface errors (e.g., grammatical gender errors in Spanish) were not treated as separate categories in the analyses, but were coded within shared higher-level dimensions (e.g., morphological errors on nouns, classified by error type), allowing for direct comparison of error distributions across Spanish and English. We have made the guide including all the possible error types and examples in English and Spanish publicly available in OSF: https://osf.io/6mwyx/?view_only=ebbfad1c0e0148c0aad2c473c4ea51f9.

The transcribers coded each error for:

1) The linguistic level where they occurred (i.e., syntactic, semantic, phonological, and morphological).

(a) We coded as syntactic errors occasions when required components of a sentence were missing (e.g., “the [dog] was barking”), duplicated e.g., “the dog dog was barking”), or produced in a wrong order (e.g., “the was dog barking”).

(b) We coded as lexico-semantic errors both when PWAs substituted a target word for a semantically related one, and when they had word finding problems that resulted in no production.

(c) Phonological and morphological errors included occasions where PWAs added, omitted, or produced an incorrect phoneme/morpheme.

(1) The class of the target word (e.g., noun, verb, pronoun, preposition, adjective, adverb).

(2) The type of error the PWAs made (i.e., addition, substitution, omission; see Figure 1 bottom right).

When a production contained multiple errors, we coded them all. For example, when a PWA produced “dauffer” for the target word “son”, it was coded as a lexico-semantic

error (as they tried to produce the semantically related word “daughter” for “son”) and a phonological error (for the additional phoneme/f/).

We also identified where production switched from one language to another (i.e., there was a code-switch) and coded for the class of the code-switched word (e.g., noun, verb . . .). When the code-switch occurred for multiple-word phrases, we indicated both that the switch happened at the phrase level and the class of the first word of the phrase.

A potential issue with using the raw number of errors/code-switches in spontaneous speech production is the considerable variation in the amount of speech produced by each PWA. It is crucial to address this because finding a mistake every 1000 words is significantly different from finding one in every other word. Therefore, to obtain a meaningful measure of PWA’s impairments, it is necessary to normalize the error rates. Two common methods of normalization in the literature involve dividing the raw number of errors by the total number of words or by the total number of utterances. For people with aphasia, the total number of utterances is often inflated by short discourse exchanges (e.g., “yeah”, “mm-hmm”), so we chose to normalize by the total number of words instead. Given that repetitions and reiterations common in PWA might also inflate the word count, we settled for normalizing the errors using the total number of words in that language excluding repetitions and retractions as the denominator. Ultimately, the three measures of discourse production (total number of utterances, and total number of words with and without retractions) were highly correlated ($r > 0.9$; $p < 0.001$), such that the choice of denominator did not have significantly impact the normalization (Figure 1(d)). We identified and excluded outliers in our normalized measures using the maximum normalized residual test (Grubbs’ test), which led to the exclusion of one participant for the code-switching analyses.

2.4. Statistical analyses

To explore the impact of premorbid language dominance on post-stroke linguistic abilities, we calculated the correlation between language dominance (as measured by PCA) and overall aphasia severity. Specifically, we calculated the correlation between language dominance and (i) the WAB aphasia severity score and (ii) the normalized amount of errors produced. Based on previous research (Kuzmina et al., 2019; Peñaloza, Barrett, et al., 2020), we expected the dominant language to show milder impairment, reflected in higher WAB scores and lower error rates.

To examine how error rates varied as a function of linguistic and lexical factors, we fitted linear mixed-effects (LME) models with normalized error rate as the dependent variable and Language Dominance (dominant vs. non-dominant), Level of Representation (phonological, morphological, syntactic, lexico-semantic), and Word Class (noun, verb, adjective, adverb, pronoun, determiner, preposition, conjunction) as fixed effects. Subject was always included as a random intercept.

Model evaluation proceeded in two steps.

- (1) Model selection: Candidate models were compared using the Bayesian Information Criterion (BIC) to identify the most parsimonious model balancing explanatory power and complexity (Burnham & Anderson, 2004). In this framework, the exclusion of fixed effects or interaction terms due to higher BIC values is interpreted as evidence that those terms are not supported by the data, and is conceptually equivalent to finding no

reliable effect for those terms (Burnham & Anderson, 2004; Wagenmakers & Farrell, 2004).

- (2) Significance testing: After identifying the optimal model, we assessed the significance of fixed effects and interactions within that model using Type III F-tests with Satterthwaite approximation for degrees of freedom (via the *lmerTest* package).

The final model minimizing the BIC included the main effects of Language Dominance, Level of Representation, and Word Class, as well as the Level of Representation \times Word Class interaction. Adding any interaction terms involving Language Dominance increased the BIC by more than 10 points, providing strong evidence against the models that included those terms.

Language identity (Spanish vs. English) was not included as a categorical predictor in the final analyses. Instead, languages were classified as dominant or non-dominant within each participant, reflecting the theoretical prediction that cross-language differences should arise as graded effects of language experience rather than as categorical typological effects. To verify that this analytic choice did not obscure language-specific patterns, we fit additional models including Language as a main effect and in interaction with Level of Representation and Word Class. Both alternatives substantially worsened model fit relative to the dominance-based model ($\Delta\text{BIC} = 11.8$ and $\Delta\text{BIC} = 120.2$, respectively), and the latter essentially accounted for all posterior model probability (BIC weight ≈ 1). Language was therefore excluded from subsequent quantitative and qualitative analyses.

Finally, we also analyzed the effect of Language Dominance on the amount of code-switching PWAs produced. To this end, we calculated the correlation between PCA language dominance and the normalized amount of code-switches. We further fitted an LME model to evaluate how Language Dominance and Word Class influence the normalized code-switching measure (random effect: subject). Both models predict a significant effect of language dominance on the amount of code-switches produced, but in opposite directions. Under an integrated language system, where lexical selection occurs on the bases of highest level of activation (Blanco-Elorrieta & Caramazza, 2021), code-switches should be more prevalent from the non-dominant to the dominant language, because items in the dominant language have higher baseline activation and are less likely to be impaired after brain damage. Contrastingly, under inhibition-based lexical selection approaches, one would expect a “reverse dominance” effect: more switches from the dominant language to the non-dominant language (Fadlon et al., 2019; Gollan et al., 2014; Li & Gollan, 2018), as a consequence of the hypothesized harder inhibition required to suppress elements from the dominant as compared to the non-dominant language.

3. Results

All effects reported below are derived from the final LME model described in Statistical analyses.

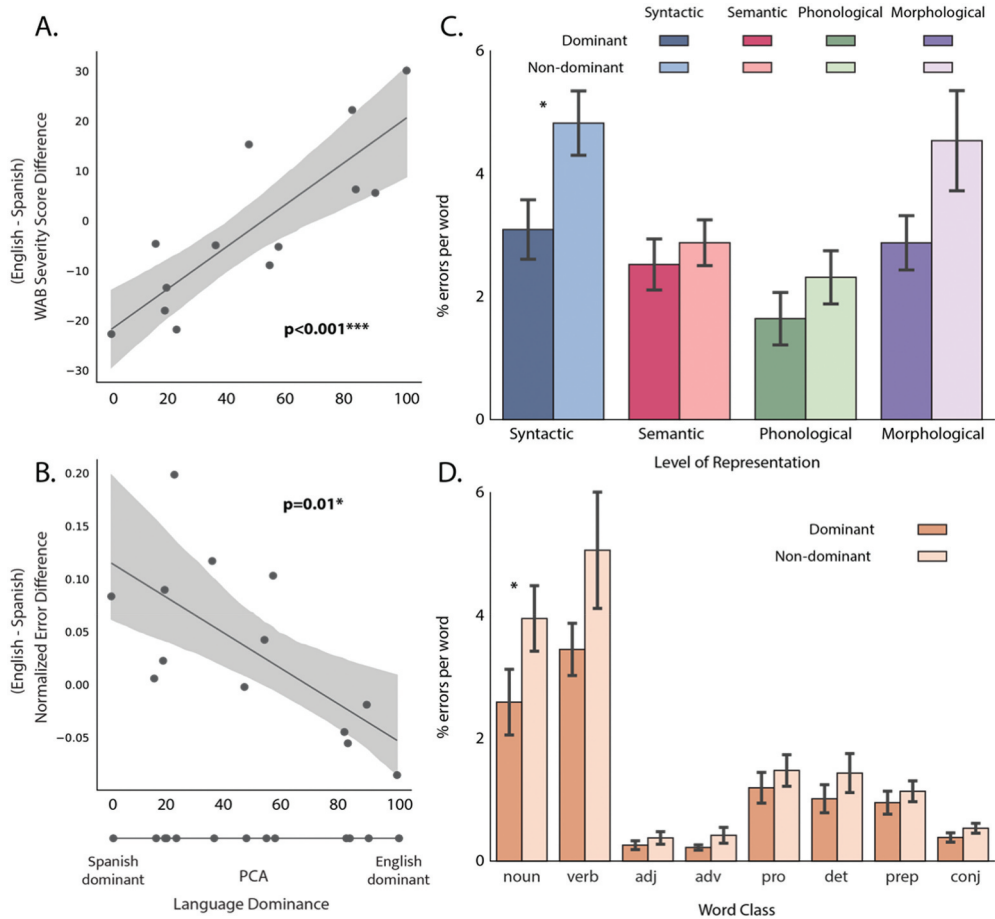


Figure 2. Effects of language dominance on post-stroke language abilities. (A) shows the relationship between language dominance and difference in WAB severity score. (B) shows the relationship between language dominance and amount of produced errors. Panels C and D show the error rates at each level of representation and word class in the dominant and non-dominant language, showing equal patterns of impairments across languages.

3.1. Uneven quantitative loss across languages: dominance predicts error quantity

PWAs' post-stroke language ability was determined by their premorbid language profile as shown by a significant correlation between the Language Dominance score (PCA) and i) the difference in the WAB score (Table 1; $r(11) = 0.83$, $p < .001$; Figure 2(a)) and ii) the amount of errors produced in the AphasiaBank tasks (Figure 2(b); $r(11) = -0.68$, $p = .01$). As predicted, the more dominant language was more spared in our bilingual PWAs. This finding was confirmed by the LME analysis, which uncovered a main effect of Language Dominance ($F(1,787) = 14.42$, $p < .001$), whereby PWAs produced more errors in their non-dominant language than in their dominant language across all levels of representation and word classes (Figures 2(c-d), 3(a-b)).

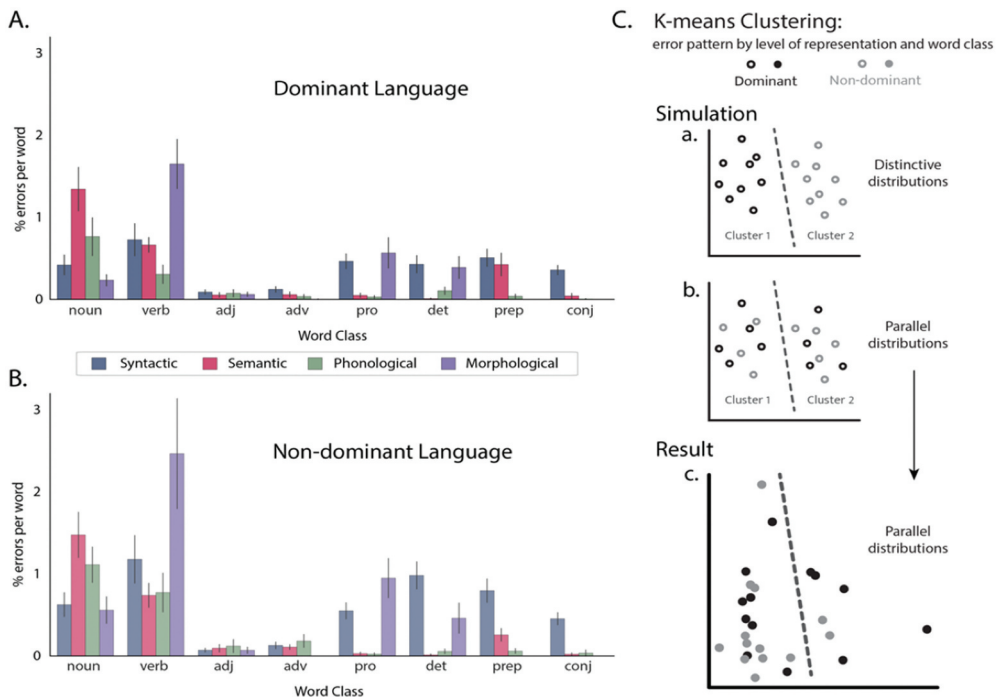


Figure 3. Effect of language dominance on error distribution. A–B: error distribution across level of representation and word class in dominant and non-dominant language. C: K-means clustering of error pattern by level of representation and word class in simulated data (a–b) and actual result (c).

3.2. Parallel qualitative loss across languages: dominance does not determine error typology

To examine whether dominant and non-dominant languages show qualitatively parallel error patterns (supporting a unified language system for bilingual individuals) or whether there are selective deficits that manifest solely in one language (revealing that languages operate on independently represented systems), we focused on evaluating the interaction terms between Language Dominance, Word Class, and Level of Representation.

The final LME model (see Statistical analyses) included the main effects of Language Dominance, Level of Representation, and Word Class, plus the Level \times Word Class interaction. Model comparison using BIC confirmed that adding either a Dominance \times Level or Dominance \times Word Class interaction worsened model fit ($\Delta\text{BIC} = +31.03$). According to conventional thresholds (Raftery, 1995), a BIC increase greater than 10 constitutes *very strong* evidence against the more complex model. In other words, the inclusion of Dominance \times Level or Dominance \times Word Class interactions did not improve the model's explanatory power sufficiently to offset their added complexity, indicating that these interactions are not supported by the data. In short, language loss shows a parallel qualitative profile across dominant and non-dominant languages (see Figure 3(a,b)).

Within the final model, we observed a main effect of Level of Representation ($F(3, 787) = 11.31, p < .001$), driven by higher error rates in morphology and syntax than in lexico-semantic or phonological levels. There was also a main effect of Word Class ($F(7, 787) =$

51.90, $p < .001$), with most errors occurring in nouns and verbs. A significant Level \times Word Class interaction ($F(21, 787) = 12.97, p < .001$) reflected that different word types afford distinct opportunities for errors (e.g., semantic errors are more frequent in content words, whereas function words are more constrained).

To further confirm that language dominance does not introduce qualitative divergence in error profiles, we conducted an unsupervised K-means clustering analysis as a complementary, data-driven test. We generated 26 vectors (one per participant per language), each containing 32 dimensions (4 representational levels \times 8 word classes), capturing the error distribution for that language. We then asked whether the algorithm could recover two clusters corresponding to dominance status. If dominance meaningfully shaped patients' impairment profile, the algorithm should separate the data into distinct clusters for dominant versus non-dominant languages; if impairment is parallel across languages, the clustering should be blind to language dominance and grouping should not be based on this dimension. This logic is illustrated in Figure 3(c), subpanels a and b, which simulate the two possible outcomes: (a) distinctive clustering by language, and (b) parallel distributions across languages.

The actual result of our clustering analysis is shown in Figure 3(c), subpanel c. The algorithm returned two clusters of similar size, but each contained a mix of dominant and non-dominant language data points (Cluster A: 8 dominant/11 non-dominant; Cluster B: 5 dominant/2 non-dominant). To quantify statistically how well the clustering aligned with the grouping expected from dominance, we used the Adjusted Rand Index (ARI). The ARI measures the similarity between two groupings on a scale from -1 to 1 , where 1 indicates a perfect match, 0 indicates similarity no better than random chance, and negative values indicate systematic disagreement. Here, we compared our grouping to the language-dominance based grouping and the observed ARI was 0.02 , effectively at chance, showing that the clustering bore no relationship to what would be expected if dominance shaped the error patterns.

To further complement the unsupervised K-means analysis, we also tested whether a supervised model (i.e., one that is explicitly told which samples are dominant vs. non-dominant) could learn to classify the error patterns by dominance. We used a support vector machine (SVM), a standard classifier in machine learning, and evaluated it using 13-fold cross-validation so that each prediction was made on data the model had not seen during training. The model's accuracy was 42% , below chance level (50%) when distinguishing two groups. This result shows that even when given the correct labels to learn from, the model could not find dominance-related structure in the data. Together, these results support that "Language Dominance" was not a meaningful dimension for grouping the data, supporting the conclusion that error distributions across languages were qualitatively parallel.

3.3. Uneven code-switches across languages: dominance predicts language switch quantity and directionality

Consistent with the prediction of an activation-based integrated model (Blanco-Elorrieta & Caramazza, 2021), we observed more code-switches from the non-dominant to the dominant language than vice versa (Figure 4). We found a significant correlation between Language Dominance and the difference in the

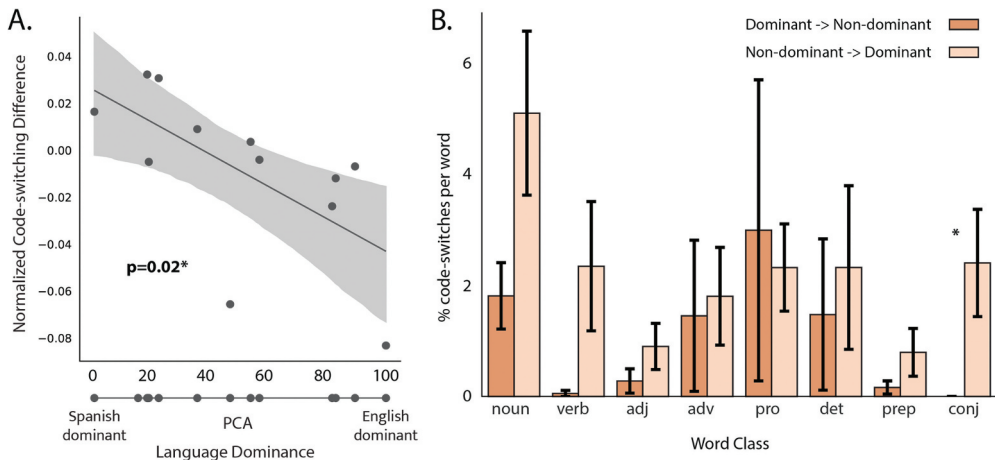


Figure 4. Effect of language dominance on the normalized code-switching measure. A: shows the correlation between PCA language dominance measure and the difference amount of code-switches. B: code-switching rate at each word class in the dominant and non-dominant language.

normalized code-switching measure ($r(10) = -0.64$, $p = .02$) and a main effect of Language Dominance in the LME model ($F(1,172) = 5.33$, $p = .02$). There was no significant main effect of Word Class ($F(7,172) = 1.85$, $p = .08$), and we did not include interaction terms as the model without them received the lowest BIC scores (-1549 vs -1518 ; $D = 31.09$), suggesting that including interaction terms did not provide a better estimation of the data over and beyond the added model complexity.

4. Discussion

4.1. Implications for bilingual language organization

The present study aimed to clarify the neurocognitive architecture of bilingual language representation by analyzing error patterns and code-switching behavior in Spanish-English bilingual individuals with post-stroke aphasia (PWA). We examined how different levels of representation (phonology, morphology, syntax, and lexico-semantics) behave in bilingual language breakdown and recovery.

A major strength of this study lies in the use of naturalistic, open-ended speech production tasks, which allowed us to evaluate bilingual language performance across all linguistic levels and word classes in ecologically valid discourse. Unlike tightly constrained experimental paradigms, narrative production affords the opportunity to capture spontaneous linguistic behavior as it unfolds in real-time, including morphological and syntactic structures that are often excluded from item-based assessments. This approach enabled us to conduct a fine-grained analysis of error types across phonology, morphology, syntax, and lexico-semantics, levels that are rarely examined simultaneously in bilingual aphasia research.

Strikingly, across participants, we found consistent qualitative profiles of impairment in both the dominant and non-dominant language not only at every level of representation but also across different word classes. This convergence across both structural levels and

lexical categories provides especially robust evidence for a shared representational system. If the two languages were supported by partially or fully distinct mechanisms, some degree of divergence would be expected in at least one linguistic domain or lexical category. Yet even when broken down into distinct word types such as content vs. function words; nouns vs. verbs vs adj etc., the same error typologies emerged in both languages, suggesting that the underlying representations and processing principles are largely shared across a bilingual's two languages, even after brain damage, although quantitative differences in performance may emerge depending on pre-morbid proficiency and usage patterns.

Importantly, this convergence should not be interpreted as surface equivalence between Spanish and English. Instead, the critical result is that the *same qualitative error structure* emerges in both the dominant and non-dominant language, indicating shared organizational structure despite quantitative asymmetries in performance.

This pattern aligns with theoretical frameworks proposing that bilinguals rely on a shared representational system (e.g., Blanco-Elorrieta & Caramazza, 2021). Under this view, when bilinguals experience language breakdown, the impairment reflects damage to a single integrated system that supports both languages, rather than to two independent language-specific systems. Consistent with this, we found that the parallelism of error types across dominant and non-dominant languages persisted even after controlling for error class, and was independently confirmed through supervised classification analysis and unsupervised clustering analysis. Although some case reports have documented dissociations across languages (e.g., Fabbro, 2001; Paradis, 2001), our findings complement rather than contradict those observations. Such dissociations may arise in single cases lesion-specific factors or from differences in experiential variables such as age of acquisition and proficiency asymmetries, which can differentially strengthen components of the shared system. At the group level, our results indicate that shared representations constitute the default organizational principle of the bilingual language system, with divergence emerging only under specific individual conditions. Crucially, the presence of individual dissociations does not challenge integrated models if those dissociations reflect uneven disruption or use-based reinforcement within a common representational architecture, rather than separate neural substrates for each language.

We also observed clear quantitative differences, with more errors in the non-dominant language. These effects are consistent with frequency-based models of language access (Gollan et al., 2014; Jacquemot et al., 2007; Martin et al., 1996), and suggest that availability rather than distinct representational formats is the primary source of asymmetry. Future studies combining behavioral data with lesion localization will be important for disambiguating these possibilities.

Turning to code-switching, we found that PWAs switched significantly more from the non-dominant to the dominant language. This pattern is consistent with activation-based models of lexical selection (Blanco-Elorrieta & Caramazza, 2021) and contrasts with inhibitory models that predict more switching into the non-dominant language (Gollan et al., 2014; Green & Abutalebi, 2013). Thus, it seems that under naturalistic, self-paced conditions, lexical selection in bilingual PWAs appears to follow patterns of relative availability, not inhibition.

Prior studies documenting reverse dominance effects often rely on forced-switch or mixed-language paradigms (Li & Gollan, 2018), which engage domain-general control

systems in a way that narrative production does not (Fedorenko & Blank, 2020). Our results instead align with recent behavioral findings in typical bilinguals and clinical populations that emphasize the communicative and facilitative role of switching (Bihovsky et al., 2023; Goral et al., 2019; Mooijman et al., 2024).

We also found a positive correlation between language proficiency and the extent of switching into the dominant language, further supporting the idea that such switches reflect facilitation. The study lacks a bilingual control group, which limits our ability to infer the aphasia-specific impact of these effects. This limitation is unavoidable, as neurologically healthy bilinguals would be unlikely to produce meaningful errors on the relatively simple set of AphasiaBank narrative tasks used in this study. Despite this constraint, our results implicate activation-based mechanisms as the primary force shaping bilingual lexical selection under impairment.

4.2. Clinical implications for bilingual aphasia

The present findings support a model in which bilingual individuals with aphasia share underlying linguistic representations across languages, even in the presence of quantitative asymmetries due to dominance. Clinically, this supports treatment approaches that target language-general mechanisms to promote cross-language transfer (Kiran & Roberts, 2010; Scimeca et al., 2023). The observed parallelism in error types across languages suggests that interventions targeting shared semantic or syntactic structures in one language may facilitate improvements in the other (Marte et al., 2025).

Code-switching behavior also has implications for diagnosis and therapy. Rather than interpreting code-switches as inherently pathological, our results indicate that bilingual PWAs may use switching as a strategy to maintain communicative effectiveness, consistent with the notion that switching can be compensatory rather than disruptive (Blanco-Elorrieta & Pykkänen, 2017; Mooijman et al., 2024). While the current study did not directly measure communicative success, we interpret switching as an adaptive behavior likely grounded in activation-based selection mechanisms. On this basis, clinicians may benefit from considering assessment protocols that allow for responses in either language and permit switching where appropriate (Goral et al., 2024). Importantly, this interpretation should not be taken as a clinical prescription. Instead, we propose that clinicians consider communicative context and underlying intent when evaluating switching behavior.

4.3. Limitations and future directions

While our sample size is modest, it reflects the recruitment challenges inherent to this population. Excluding participants with limited discourse output introduces selection bias, as these individuals likely had more severe aphasia and/or significant disparities in proficiency. This exclusion may limit the generalizability of our findings to the full spectrum of bilingual aphasia severity. Longitudinal studies and larger samples are essential for evaluating whether recovery trajectories differ across levels of representation and between languages. Such work will be critical for refining theories of bilingual language organization and informing personalized, evidence-based interventions.

5. Conclusion

This study provides empirical support for a model of bilingual language organization in aphasia that relies on shared neural and cognitive resources across languages, modulated by relative language dominance. Parallel qualitative patterns of impairment across languages, together with activation-based code-switching profiles, suggest that bilinguals rely on an integrated language system that spans multiple levels of representation. These findings have both theoretical and clinical implications, advocating for models of bilingualism that incorporate shared architecture alongside usage-dependent variation, and for intervention strategies that leverage cross-language commonalities to support recovery in bilingual aphasia. Because neural injury reveals rather than creates underlying structures, these findings also support the inference that shared representations were present in the neurotypical brain prior to injury. However, direct comparisons with healthy bilinguals will be critical to fully establish which aspects of this integration are universal and which may be shaped by post-stroke adaptation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by NIH grants [R00 DC019973-03] to E.B.E., [R01DC020965-02S1] to MM and NIH [R01DC020653] to S.K. [U01 DC014922] to S.K. and M.J.M. was partially supported by [T32 DC013017].

ORCID

Esti Blanco-Elorrieta  <http://orcid.org/0000-0003-2983-5880>

Data statement

The data will be publicly available in a OSF repository.

References

- Ansaldo, A. I., & Saidi, L. G. (2014). Aphasia therapy in the age of globalization: Cross-linguistic therapy effects in bilingual aphasia. *Behavioural Neurology*, 2014(1), 603085. <https://doi.org/10.1155/2014/603085>
- Bastiaanse, R., & Thompson, C. K. (Eds.). (2012). *Perspectives on agrammatism*. Psychology Press.
- Bihovsky, A., Ben-Shachar, M., & Meir, N. (2023). Language abilities, not cognitive control, predict language mixing behavior in bilingual speakers with aphasia. *Journal of Communication Disorders*, 105, 106367. <https://doi.org/10.1016/j.jcomdis.2023.106367>
- Biondo, N., Molinaro, N., & Mancini, S. (2023). The neurolinguistics of the second language morphological system: The role of grammar-related and speaker-related factors. In *The Routledge handbook of second language acquisition and neurolinguistics* (pp. 115–132). Routledge.

- Blanco-Elorrieta, E., & Caramazza, A. (2021). A common selection mechanism at each linguistic level in bilingual and monolingual language production. *Cognition*, 213, 104625. <https://doi.org/10.1016/j.cognition.2021.104625>
- Blanco-Elorrieta, E., & Pyllkkänen, L. (2017). Bilingual language switching in the laboratory versus in the wild: The spatiotemporal dynamics of adaptive language control. *Journal of Neuroscience*, 37(37), 9022–9036. <https://doi.org/10.1523/JNEUROSCI.0553-17.2017>
- Brookshire, R. H., & Nicholas, L. E. (1994). Speech sample size and test-retest stability of connected speech measures for adults with aphasia. *Journal of Speech, Language, and Hearing Research*, 37(2), 399–407. <https://doi.org/10.1044/jshr.3702.399>
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, 3(4), 572–582. [https://doi.org/10.1016/0093-934X\(76\)90048-1](https://doi.org/10.1016/0093-934X(76)90048-1)
- Cargnelutti, E., Tomasino, B., & Fabbro, F. (2019). Language brain representation in bilinguals with different age of appropriation and proficiency of the second language: A meta-analysis of functional imaging studies. *Frontiers in Human Neuroscience*, 13, 445658. <https://doi.org/10.3389/fnhum.2019.00154>
- Chee, M. W., Caplan, D., Soon, C. S., Sriram, N., Tan, E. W., Thiel, T., & Weekes, B. (1999). Processing of visually presented sentences in mandarin and English studied with fMRI. *Neuron*, 23(1), 127–137. [https://doi.org/10.1016/S0896-6273\(00\)80759-X](https://doi.org/10.1016/S0896-6273(00)80759-X)
- Chen, C., Gong, X. L., Tseng, C., Klein, D. L., Gallant, J. L., & Deniz, F. (2024). Bilingual language processing relies on shared semantic representations that are modulated by each language. *BioRxiv*, 2024–2026.
- Declerck, M., Kleinman, D., & Gollan, T. H. (2020). Which bilinguals reverse language dominance and why? *Cognition*, 204, 104384. <https://doi.org/10.1016/j.cognition.2020.104384>
- Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D., Van de Moortele, P.-F., Lehericy, S., & Le Bihan, D. (1997). Anatomical variability in the cortical representation of first and second language. *Neuroreport*, 8(17), 3809–3815. <https://doi.org/10.1097/00001756-199712010-00030>
- Fabbro, F. (2001). The bilingual brain: Bilingual aphasia. *Brain and Language*, 79(2), 201–210. <https://doi.org/10.1006/brln.2001.2480>
- Fadlon, J., Li, C., Prior, A., & Gollan, T. H. (2019). Using what's there: Bilinguals adaptively rely on orthographic and color cues to achieve language control. *Cognition*, 191, 103990. <https://doi.org/10.1016/j.cognition.2019.06.002>
- Fedorenko, E., & Blank, I. A. (2020). Broca's area is not a natural kind. *Trends in Cognitive Sciences*, 24(4), 270–284. <https://doi.org/10.1016/j.tics.2020.01.001>
- Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, 91(4), 1357–1392. <https://doi.org/10.1152/physrev.00006.2011>
- Gao, F., Hua, L., Armada-da-Silva, P., Zhang, J., Li, D., Chen, Z., Yuan, Z., Du, M., & Yuan, Z. (2023). Shared and distinct neural correlates of first and second language morphological processing in bilingual brain. *NPJ Science of Learning*, 8(1), 33. <https://doi.org/10.1038/s41539-023-00184-9>
- Ghazi-Saidi, L., & Ansaldi, A. I. (2017). The neural correlates of semantic and phonological transfer effects: Language distance matters. *Bilingualism: Language and Cognition*, 20(5), 1080–1094. <https://doi.org/10.1017/S136672891600064X>
- Goldrick, M., & Gollan, T. H. (2023). Inhibitory control of the dominant language: Reversed language dominance is the tip of the iceberg. *Journal of Memory and Language*, 130, 104410. <https://doi.org/10.1016/j.jml.2023.104410>
- Golestani, N., Alario, F. X., Meriaux, S., Le Bihan, D., Dehaene, S., & Pallier, C. (2006). Syntax production in bilinguals. *Neuropsychologia*, 44(7), 1029–1040. <https://doi.org/10.1016/j.neuropsychologia.2005.11.009>

- Gollan, T. H., & Ferreira, V. S. (2009). Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 640. <https://doi.org/10.1037/a0014981>
- Gollan, T. H., Schotter, E. R., Gomez, J., Murillo, M., & Rayner, K. (2014). Multiple levels of bilingual language control: Evidence from language intrusions in reading aloud. *Psychological Science*, 25(2), 585–595. <https://doi.org/10.1177/0956797613512661>
- Goral, M., Antolovic, K., Hejazi, Z., & Schulz, F. M. (2024). Using a translanguaging framework to examine language production in a trilingual person with aphasia. *Clinical Linguistics & Phonetics*, 39(1), 1–20. <https://doi.org/10.1080/02699206.2024.2328240>
- 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. (2005). The neurocognition of recovery patterns in bilingual aphasics. *Handbook of Bilingualism: Psycholinguistic Approaches*, 516–530.
- 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>
- Green, D. W., Grogan, A., Crinion, J., Ali, N., Sutton, C., & Price, C. J. (2010). Language control and parallel recovery of language in individuals with aphasia. *Aphasiology*, 24(2), 188–209. <https://doi.org/10.1080/02687030902958316>
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9(9), 416–423. <https://doi.org/10.1016/j.tics.2005.07.004>
- Hasegawa, M., Carpenter, P. A., & Just, M. A. (2002). An fMRI study of bilingual sentence comprehension and workload. *Neuroimage*, 15(3), 647–660. <https://doi.org/10.1006/nimg.2001.1001>
- Ip, K. I., Hsu, L. S. J., Arredondo, M. M., Tardif, T., & Kovelman, I. (2017). Brain bases of morphological processing in Chinese-English bilingual children. *Developmental Science*, 20(5), e12449. <https://doi.org/10.1111/desc.12449>
- Jacquemot, C., Dupoux, E., & Bachoud-Lévi, A. C. (2007). Breaking the mirror: Asymmetrical disconnection between the phonological input and output codes. *Cognitive Neuropsychology*, 24(1), 3–22. <https://doi.org/10.1080/02643290600683342>
- Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: A case-series comparison. *Brain*, 129(Pt 8), 2132–2147. <https://doi.org/10.1093/brain/awl153>
- Kastenbaum, J. G., Bedore, L. M., Peña, E. D., Sheng, L., Mavis, I., Sebastian-Vaytadden, R., Rangamani, G., Vallila-Rohter, S., & Kiran, S. (2019). The influence of proficiency and language combination on bilingual lexical access. *Bilingualism: Language and Cognition*, 22(2), 300–330. <https://doi.org/10.1017/S1366728918000366>
- Kertesz, A. (2007). *Western aphasia battery—revised*. The Psychological Corporation.
- Kiran, S., Grasemann, U., Sandberg, C., & Miikkulainen, R. (2013). A computational account of bilingual aphasia rehabilitation. *Bilingualism: Language and Cognition*, 16(2), 325–342. <https://doi.org/10.1017/S1366728912000533>
- Kiran, S., & Iakupova, R. (2011). Understanding the relationship between language proficiency, language impairment and rehabilitation: Evidence from a case study. *Clinical Linguistics & Phonetics*, 25(6–7), 565–583. <https://doi.org/10.3109/02699206.2011.566664>
- Kiran, S., & Roberts, P. M. (2010). Semantic feature analysis treatment in Spanish-English and French-English bilingual aphasia. *Aphasiology*, 24(2), 231–261. <https://doi.org/10.1080/02687030902958365>
- Kleinman, D., & Gollan, T. H. (2018). Inhibition accumulates over time at multiple processing levels in bilingual language control. *Cognition*, 173, 115–132. <https://doi.org/10.1016/j.cognition.2018.01.009>

- 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>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kuzmina, E., Goral, M., Norvik, M., & Weekes, B. S. (2019). What influences language impairment in bilingual aphasia? A meta-analytic review. *Frontiers in Psychology*, 10, 435219. <https://doi.org/10.3389/fpsyg.2019.00445>
- Lallini, N., Miller, N., & Howard, D. (2007). Lexical influences on single word repetition in acquired spoken output impairment: A cross language comparison. *Aphasiology*, 21(6–8), 617–631. <https://doi.org/10.1080/02687030701192067>
- Lehtonen, M., Vorobyev, V., Soveri, A., Hugdahl, K., Tuokkola, T., & Laine, M. (2009). Language-specific activations in the brain: Evidence from inflectional processing in bilinguals. *Journal of Neurolinguistics*, 22(5), 495–513. <https://doi.org/10.1016/j.jneuroling.2009.05.001>
- Lerman, A., Bihovsky, A., Meir, N., Ben-Shachar, M., & Goral, M. (2025). Language mixing in people with aphasia: A cross-linguistic analysis using the 4M model. *Brain and Language*, 269, 105616. <https://doi.org/10.1016/j.bandl.2025.105616>
- Lerman, A., Edmonds, L. A., & Goral, M. (2019). Cross-language generalisation in bilingual aphasia: What are we missing when we do not analyse discourse? *Aphasiology*, 33(9), 1154–1162. <https://doi.org/10.1080/02687038.2018.1538493>
- Lerman, A., Goral, M., & Obler, L. K. (2020). The complex relationship between pre-stroke and post-stroke language abilities in multilingual individuals with aphasia. *Aphasiology*, 34(11), 1319–1340. <https://doi.org/10.1080/02687038.2019.1673303>
- Li, C., & Gollan, T. H. (2018). Cognates interfere with language selection but enhance monitoring in connected speech. *Memory & Cognition*, 46(6), 923–939. <https://doi.org/10.3758/s13421-018-0812-x>
- MacWhinney, B. (2017). *Tools for analyzing talk part 2: The CLAN program*. Talkbank. org, (2000).
- MacWhinney, B. (2022). The competition model: Past and future. *A Life in Cognition: Studies in Cognitive Science in Honor of Csaba Pléh* (pp. 3–16). Springer International Publishing.
- MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, 25(11), 1286–1307. <https://doi.org/10.1080/02687038.2011.589893>
- MacWhinney, B., Kroll, J. F., & De Groot, A. M. B. (2005). A unified model of language acquisition. *Handbook of Bilingualism: Psycholinguistic Approaches*, 4967, 50–70.
- MacWhinney, B., & Wagner, J. (2010). Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository. *Gesprächsforschung: Online-Zeitschrift Zur Verbalen Interaktion*, 11, 154.
- Malik-Moraleda, S., Jouravlev, O., Taliaferro, M., Mineroff, Z., Cucu, T., Mahowald, K., Fedorenko, E., & Fedorenko, E. (2024). Functional characterization of the language network of polyglots and hyperpolyglots with precision fMRI. *Cerebral Cortex*, 34(3), bhae049. <https://doi.org/10.1093/cercor/bhae049>
- Marte, M. J., Carpenter, E., Falconer, I. B., Scimeca, M., Abdollahi, F., Peñaloza, C., & Kiran, S. (2022). LEX-BADAT: Language experience in bilinguals with and without aphasia dataset. *Frontiers in Psychology*, 13, 875928.
- Marte, M. J., Carpenter, E., Scimeca, M., Russell-Meill, M., Peñaloza, C., Grasemann, U., Kiran, S., & Kiran, S. (2025). Machine learning predictions of recovery in bilingual poststroke aphasia: Aligning insights with clinical evidence. *Stroke*, 56(2), 494–504. <https://doi.org/10.1161/STROKEAHA.124.047867>
- Martin, N., Saffran, E. M., & Dell, G. S. (1996). Recovery in deep dysphasia: Evidence for a relation between auditory-verbal STM capacity and lexical errors in repetition. *Brain and Language*, 52(1), 83–113. <https://doi.org/10.1006/brln.1996.0005>
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88(5), 375. <https://doi.org/10.1037/0033-295X.88.5.375>

- Meykadeh, A., Golfam, A., Batouli, S. A. H., & Sommer, W. (2021). Overlapping but language-specific mechanisms in morphosyntactic processing in highly competent L2 acquired at school entry: fMRI evidence from an alternating language switching task. *Frontiers in Human Neuroscience*, *15*, 728549. <https://doi.org/10.3389/fnhum.2021.728549>
- Molinaro, N., Nara, S., & Carreiras, M. (2024). Early language dissociation in bilingual minds: Magnetoencephalography evidence through a machine learning approach. *Cerebral Cortex*, *34*(2), bhae053. <https://doi.org/10.1093/cercor/bhae053>
- Mooijman, S., Schoonen, R., Ruiters, A., Roelofs, M. B., & Ruiters, M. B. (2024). Benefits of free language choice in bilingual individuals with aphasia. *Aphasiology*, 1–39. <https://doi.org/10.1080/02687038.2025.2452928>
- Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., & Weiller, C. (2003). Broca's area and the language instinct. *Nature Neuroscience*, *6*(7), 774–781. <https://doi.org/10.1038/nn1077>
- Nadeau, S. E. (2019). Bilingual aphasia: Explanations in population encoding. *Journal of Neurolinguistics*, *49*, 117–143. <https://doi.org/10.1016/j.jneuroling.2018.10.002>
- Ou, J., Li, W., Yang, Y., Wang, N., & Xu, M. (2020). Earlier second language acquisition is associated with greater neural pattern dissimilarity between the first and second languages. *Brain and Language*, *203*, 104740. <https://doi.org/10.1016/j.bandl.2019.104740>
- Paradis, M. (2001). Bilingual and polyglot aphasia. In R. S. Berndt (Ed.), *Handbook of neuropsychology* (2nd ed. pp. 69–91). Elsevier Science.
- Paradis, M. (2004). A neurolinguistic theory of bilingualism.
- Peñalosa, C., Barrett, K., & Kiran, S. (2020). The influence of prestroke proficiency on poststroke lexical-semantic performance in bilingual aphasia. *Aphasiology*, *34*(10), 1223–1240. <https://doi.org/10.1080/02687038.2019.1666082>
- Peñalosa, C., Dekhtyar, M., Scimeca, M., Carpenter, E., Mukadam, N., & Kiran, S. (2020). Predicting treatment outcomes for bilinguals with aphasia using computational modeling: Study protocol for the PROCoM randomised controlled trial. *BMJ Open*, *10*(11), e040495. <https://doi.org/10.1136/bmjopen-2020-040495>
- Peñalosa, C., Grasmann, U., Dekhtyar, M., Miikkulainen, R., & Kiran, S. (2019). BiLex: A computational approach to the effects of age of acquisition and language exposure on bilingual lexical access. *Brain and Language*, *195*, 104643. <https://doi.org/10.1016/j.bandl.2019.104643>
- Peñalosa, C., & Kiran, S. (2019). Recovery and rehabilitation patterns in bilingual and multilingual aphasia. *The Handbook of the Neuroscience of Multilingualism*, 553–571.
- Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, *15*(2), 202–206. <https://doi.org/10.1016/j.conb.2005.03.007>
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111–163.
- Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*(1), 60. <https://doi.org/10.1037/0033-295X.89.1.60>
- Sakai, K. L., Miura, K., Narafu, N., & Muraishi, Y. (2004). Correlated functional changes of the prefrontal cortex in twins induced by classroom education of second language. *Cerebral Cortex*, *14*(11), 1233–1239.
- Scimeca, M., Peñalosa, C., & Kiran, S. (2023). Multilevel factors predict treatment response following semantic feature-based intervention in bilingual aphasia. *Bilingualism: Language and Cognition*, *27*(2), 1–17. <https://doi.org/10.1017/S1366728923000391>
- Singleton, D. (2005). The critical period hypothesis: A coat of many colours. *IRAL - International Review of Applied Linguistics in Language Teaching*, *43*(4), 269–285. <https://doi.org/10.1515/iral.2005.43.4.269>
- Singleton, D., & Leńniewska, J. (2021). The critical period hypothesis for L2 acquisition: An unfalsifiable embarrassment? *Languages*, *6*(3), 149. <https://doi.org/10.3390/languages6030149>
- Tettamanti, M., Alkadhi, H., Moro, A., Perani, D., Kollias, S., & Weniger, D. (2002). Neural correlates for the acquisition of natural language syntax. *Neuroimage*, *17*(2), 700–709. <https://doi.org/10.1006/nimg.2002.1201>

- Tyler, L. K., Marslen-Wilson, W. D., Randall, B., Wright, P., Devereux, B. J., Zhuang, J., Stamatakis, E. A., & Stamatakis, E. A. (2011). Left inferior frontal cortex and syntax: Function, structure and behaviour in patients with left hemisphere damage. *Brain*, *134*(2), 415–431. <https://doi.org/10.1093/brain/awq369>
- Van de Putte, E., De Baene, W., Price, C. J., & Duyck, W. (2018). Neural overlap of L1 and L2 semantic representations across visual and auditory modalities: A decoding approach. *Neuropsychologia*, *113*, 68–77. <https://doi.org/10.1016/j.neuropsychologia.2018.03.037>
- Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. <https://doi.org/10.3758/BF03206482>
- Yang, Y., Wang, J., Bailer, C., Cherkassky, V., & Just, M. A. (2017). Commonality of neural representations of sentences across languages: Predicting brain activation during Portuguese sentence comprehension using an English-based model of brain function. *Neuroimage*, *146*, 658–666. <https://doi.org/10.1016/j.neuroimage.2016.10.029>