



Multilevel factors predict treatment response following semantic feature-based intervention in bilingual aphasia

Michael Scimeca¹ , Claudia Peñaloza^{2,3,4} and Swathi Kiran¹

Research Article

Cite this article: Scimeca, M., Peñaloza, C., & Kiran, S. (2024). Multilevel factors predict treatment response following semantic feature-based intervention in bilingual aphasia. *Bilingualism: Language and Cognition*, 27, 246–262. <https://doi.org/10.1017/S1366728923000391>

Received: 15 June 2022
Revised: 17 April 2023
Accepted: 18 April 2023
First published online: 22 August 2023

Keywords:

bilingual aphasia; anomia therapy; semanticfeature treatment; treatment outcomes; cross-language generalization

Corresponding author:

Michael Scimeca
Aphasia Research Laboratory
Dept. of Speech, Language, and Hearing Sciences
College of Health & Rehabilitation Sciences:
Sargent College
635 Commonwealth Ave.
Boston, MA, USA 02215
Email: mscimeca@bu.edu

¹Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA, USA; ²Department of Cognition, Development and Educational Psychology, Faculty of Psychology, University of Barcelona, Barcelona, Spain; ³Institute of Neurosciences, University of Barcelona, Barcelona, Spain and ⁴Cognition and Brain Plasticity Unit, Bellvitge Biomedical Research Institute-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain

Abstract

Semantic feature-based treatments (SFTs) are effective rehabilitation strategies for word retrieval deficits in bilinguals with aphasia (BWA). However, few studies have prospectively evaluated the effects of key parameters of these interventions on treatment outcomes. This study examined the influence of intervention-level (i.e., treatment language and treatment sessions), individual-level (baseline naming severity and age), and stimulus-level (i.e., lexical frequency, phonological length, and phonological neighborhood density) factors on naming improvement in a treated and untreated language for 34 Spanish–English BWA who completed 40 hours of SFT. Results revealed significant improvement over time in both languages. In the treated language, individuals who received therapy in their L1 improved more. Additionally, higher pre-treatment naming scores predicted greater response to treatment. Finally, a frequency effect on baseline naming accuracy and phonological effects on accuracy over time were associated with differential treatment gains. These findings indicate that multi-level factors are influential predictors of bilingual treatment outcomes.

1. Introduction

Language interventions are essential to improve the communication deficits that impact the quality of life of bilinguals with aphasia (BWA). Following an acquired brain injury (ABI), BWA often present to clinical programs with heterogeneous patterns of impairment which arise due to the complex interactions between lesion and pre-and post-morbid language proficiency characteristics (Fabbro, 2001; Paradis, 2004; Peñaloza et al., 2019, 2020a; also see Kuzmina et al., 2019; Lorenzen & Murray, 2008 for a review). Previous bilingual aphasia rehabilitation research has mainly focused on the efficacy of language treatment and the effects of bilingualism-related variables on response to treatment in this population. Studies addressing these questions have provided some evidence for positive treatment outcomes in particular in the treated language (see Peñaloza & Kiran, 2019; also see Faroqi-Shah et al., 2010 for a review) and evidence for modulatory effects of language proficiency on overall response to treatment (see Goral & Lerman, 2020 for a review).

It should be noted however, that improvement following aphasia rehabilitation is also highly variable among BWA. This variation in treatment outcomes (see Goral & Lerman, 2020 for a review) may be partially attributed to different methodological approaches used in bilingual rehabilitation research aiming to address a variety of clinical questions which remain unresolved (Peñaloza & Kiran, 2019). Some of these open questions concern INTERVENTION-LEVEL FACTORS such as which language should be selected for therapy. A second set of questions is related to INDIVIDUAL-LEVEL FACTORS such as whether post-stroke aphasia severity, demographic characteristics (e.g., age, education, etc.), and bilingualism-related characteristics (e.g., language proficiency) may predict participant-specific rates of language improvement. Other questions seek to understand whether STIMULUS-LEVEL FACTORS (i.e., training specific stimuli) such as high frequency words may affect therapy outcomes. Some studies have addressed these questions, but small sample sizes and a large variability in the language interventions reported across studies have constrained our ability to draw conclusions about optimal treatment design and delivery (Peñaloza & Kiran, 2019). Establishing appropriate guidelines for clinical services will be even more crucial in the coming years given that higher life expectancy, general trends towards multiculturalism within communities, and rising stroke incidence among racial-ethnic minority groups (Rodriguez et al., 2014) signal that bilingual individuals will experience more ABIs and will constitute an ever-growing proportion of adult speech and language caseloads (Centeno, 2008, 2009).

Given the need to inform clinical guidelines for bilingual intervention, the present study focused on evaluating the effects of a pre-defined set of intervention-level (treatment language

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



and number of treatment sessions completed), individual-level (baseline naming severity and age), and stimulus-level (lexical frequency, phonological length, and phonological neighborhood density) parameters on anomia treatment outcomes in BWA above and beyond previous research which has mainly focused on examining the relationship between bilingualism-related variables and treatment outcomes (Goral & Lerman, 2020). In the following sections, we first provide a review of factors that influence treatment response in this population and then further define our own factors of interest within a multilevel framework.

1.1. Intervention-level factors modulating language treatment outcomes in BWA

Intervention-level factors which may influence rehabilitation outcomes in bilingual aphasia include treatment type, treatment language (e.g., L1 or the first acquired language versus L2 or the second acquired language), and time in treatment.

Treatment type

Previous studies have generally shown that language recovery in BWA is possible across a variety of intervention approaches targeting specific language deficits in L1 or L2 (Ansaldo & Saidi, 2014; Goral & Lerman, 2020; Faroqi-Shah et al., 2010; Kohnert, 2009; Peñaloza & Kiran, 2019).

As anomia is considered the hallmark residual deficit across aphasia syndromes (Laine & Martin, 2006), most studies have targeted word retrieval deficits and have shown evidence for naming improvement following therapy (Abutalebi et al., 2009; Ansaldo et al., 2010; Croft et al., 2011; Edmonds & Kiran, 2006; Galvez & Hinckley, 2003; Gil & Goral, 2004; Hinckley, 2003; Hughes et al., 2012; Keane & Kiran, 2015; Kiran et al., 2013; Kiran & Roberts, 2010; Kiran & Iakupova, 2011; Kohnert, 2004; Laganaro et al., 2003; Li et al., 2020; Maragnolo et al., 2009; Meinzer et al., 2007; Peñaloza et al., 2021; Roberts & Tainturier, 2010; Sandberg et al., 2021). Importantly, semantic feature-based treatment (SFT) for word retrieval difficulties has been associated with not only significant treatment gains in the treated language but also a moderate degree of cross-language generalization (Croft et al., 2011; Edmonds & Kiran, 2006; Kiran & Roberts, 2010; Kiran et al., 2013; Kiran & Iakupova, 2011; Peñaloza et al., 2021). Spreading activation has been proposed as a potential mechanism to account for cross-language generalization effects in BWA following SFT as this treatment capitalizes on a well-known framework of bilingual lexical processing which assumes (i) a semantic system with shared conceptual representations for words across languages while lexical representations are independent for each language (Kroll & Stewart, 1994) and (ii) activation propagating from the semantic to the lexical system in a language independent manner (Costa et al., 2006).

Thus, training the semantic features of a given word is assumed to strengthen its corresponding representation in the mental lexicon at the semantic-conceptual level that then propagates to the corresponding lexical-phonetic representations making it available for production (Goral & Lerman, 2020). In this way, increased activation can be specific, accounting for the improvements restricted to just the treated language, or it can spread to the untreated language accounting for cross-language generalization effects (Kiran et al., 2013). Considering theoretical models of the bilingual mental lexicon (Kroll & Stewart, 1994) and as recently shown via computational modeling (Grasemann et al., 2021) cross-language generalization could be achieved via

associative connections between the semantic and the untreated language lexical systems or via associative connections between the two lexical systems.

Treatment language

Another factor that may be crucial for treatment outcome predictions is the language selected for therapy. As noted in the previous section, treatment in either L1 or L2 may lead to improvement in that language, but some evidence suggests that gains in the treated language may be larger for BWA receiving therapy in their L1 (see Faroqi-Shah et al., 2010 for a review) relative to therapy in their L2 (i.e., treatment delivered in language blocks; Gil & Goral, 2004) and when compared to other participant groups receiving therapy in their L2 (Croft et al., 2011; Junqué et al., 1989; Kiran et al., 2013).

Additionally, selecting a specific language for therapy may determine the degree of generalization to the untreated language. In balanced BWA, cross-language generalization is often observed across languages that show comparable ability (Edmonds & Kiran, 2006; Maragnolo et al., 2009), suggesting that targeting either language may be similarly beneficial. However, cross-language transfer effects may be more sensitive to the language chosen for treatment in unbalanced BWA. For instance, Edmonds and Kiran (2006) and Kiran and Iakupova (2011) have shown that treatment in L2 resulted in greater cross-language generalization to L1 given hypothesized stronger connections from the L2 to L1 lexicon (Kroll & Stewart, 1994). However, evidence has emerged which opposes this theory. For example, Abutalebi et al. (2009) observed antagonistic recovery, characterized by reduced performance in L1 Spanish at the end of treatment, for a Spanish-Italian bilingual who completed language therapy in their L2 Italian. Other studies have found that treatment in L1 generalized more to untreated L2 (Croft et al., 2011; Galvez & Hinckley, 2003; Gil & Goral, 2004) or have demonstrated no generalization to L1 following treatment in L2 (Kiran & Roberts, 2010; Meinzer et al., 2007; Miller Amberber, 2012).

Clearly, selecting which language to target in treatment has implications for both the degree of improvement in the treated language and the amount of cross-language generalization in the untreated language. These mixed results may be partially explained by factors such as language impairment and pre-stroke language proficiency which may alter the associative connections between the separate lexicons and therefore disrupt expected patterns of lexical access (Goral & Lerman, 2020; Peñaloza & Kiran, 2019). Leveraging these factors to determine which language to administer therapy is a subject of ongoing debate in bilingual aphasia rehabilitation and deserves more in-depth consideration in the literature.

Time in treatment

Measures of time spent in therapy may also influence response to treatment. Many studies have reported positive treatment outcomes at different dosages at the end of treatment such as Croft et al. (2011) and Kiran et al. (2013) who administered 20 hours and on average 40 hours of therapy, respectively. Although findings from studies which report longitudinal measures of recovery (Croft et al., 2011; Edmonds & Kiran, 2006; Kiran & Roberts, 2010) generally suggest that more therapy would be beneficial, the granular effect of time in treatment (i.e., number of treatment sessions completed) has only been investigated recently (Li et al., 2020). Nevertheless, more work is needed with larger participant groups to illuminate the importance of this factor in predicting rehabilitation outcomes.

1.2. Individual-level factors modulating language treatment outcomes in BWA

A second set of factors that may modulate treatment outcomes concerns individual BWA characteristics. At the individual-level, many bilingualism-related variables such as language proficiency, language dominance, and linguistic distance have been investigated. We refer the reader to Goral and Lerman (2020) which provides an extensive review on such factors. In this section, we instead focus on post-stroke language impairment and demographic characteristics whose possible effects on treatment outcomes have not been well studied thus far and are central to the present study.

Language impairment

Thus far, most studies of bilingual aphasia rehabilitation that have examined language impairment have used clinical assessment scores to compare patterns of deficits across a bilingual's two or more languages after ABI (see Kuzmina et al., 2019 for a review). Generally, overall aphasia severity and the severity of the specific language impairment being targeted in treatment (e.g., word retrieval deficits) as measured by standardized language tests in each language have emerged as important predictors of post-stroke aphasia recovery (Lazar et al., 2010; Plowman et al., 2012; Watila & Balarabe, 2015; also see Kristinsson et al., 2022 for a review). Importantly, one recent study demonstrated that pre-treatment naming severity predicts treatment gains following SFT for word retrieval deficits in monolingual aphasia (Quique et al., 2019). These findings suggest that prospective analyses of language impairment variables and their effects on treatment outcomes may be useful in predicting recovery in treated and untreated languages in bilingual aphasia, though they have not been previously examined.

Demographic characteristics

Demographic variables such as age, education, and time post-onset may also be important predictors of treatment outcomes. Although some studies with monolinguals with aphasia have reported associations between education and aphasia severity (see Kristinsson et al., 2022 for a review), it remains unknown whether education modulates the amount of post-treatment language improvement. Additionally, while time post-ABI onset is a useful predictor of early language recovery, it sharply declines in importance during chronic recovery from aphasia and in some cases, it may not predict recovery at all (Moss & Nicholas, 2006; Nardo et al., 2017).

Unlike education and time-post onset, age has emerged as a more consistent predictor of treatment outcomes in studies involving monolingual speakers in which greater post-treatment improvement has been observed in younger individuals relative to older individuals (Nakagawa et al., 2019). These patterns may be partially attributed to declining brain plasticity in typical aging (Shafto & Tyler, 2014) which may limit the amount of observable language recovery as a function of age. Critically, these factors have not been examined in response to treatment for BWA.

1.3. Stimulus-level factors modulating language treatment outcomes in BWA

Although less frequently studied in the language rehabilitation literature, stimulus-level factors such as the psycholinguistic

properties of words chosen for treatment may also modulate treatment response. In monolingual aphasia, it has been established that lexical frequency modulates naming abilities (Kittredge et al., 2008; Nozari et al., 2010) – such that words which occur more often in everyday experiences tend to be retrieved faster and more accurately. Shorter word length (Castro et al., 2020; Nickels & Howard, 1995, 2004) and denser phonological neighborhoods (Gordon, 2002; Gordon & Dell, 2001; Laganaro et al., 2006; Middleton & Schwartz, 2011) have also emerged as facilitators of word retrieval in aphasia. Though these effects have been examined outside the context of treatment, one study (Hendricks et al., 2014) recently showed that words with high phonological neighborhood density benefitted the most during a phonological treatment when compared to low density words and phonological control words. Nevertheless, no study has systematically examined how these psycholinguistic properties might influence naming abilities and response to treatment in bilingual aphasia.

1.4. The present study

Although studies in bilingual aphasia have begun to examine intervention-, individual-, and stimulus-level factors as described earlier in the introduction, few have prospectively defined a set of variables for use in predicting treatment outcomes in BWA. Furthermore, none has examined the effects of all three factor levels (i) on a single set of patient outcomes observed in the treated and the untreated language or (ii) throughout treatment using fine-grained, time-series analyses.

In the current study, we systematically examined the predictive capacity of a set of multilevel factors on treatment outcomes in the treated and the untreated language for a large cohort of 34 Spanish–English BWA who had received SFT in either L1 (i.e., their first acquired language) or L2 (i.e., their second acquired language). The examination of such factors in the context of SFT is highly relevant, since SFT has been shown to be an effective intervention for word retrieval deficits (Edmonds & Kiran, 2006; Kiran & Roberts, 2010; Kiran et al., 2013; Peñaloza et al., 2021) in bilingual aphasia. We defined our research aims and hypotheses as follows:

- 1) To what extent do intervention-level factors predict naming outcomes in the treated and untreated language? To address this question, we used the number of treatment sessions completed and treatment language (i.e., L1 or L2) to compare accuracy for trained items relative to control words in each language. We expected that accuracy for trained items in the treated language and their corresponding translations in the untreated language would be higher than that of control items in both languages over time. Additionally, in line with previous observations, we expected BWA who received therapy in their L1 to show greater within-language naming improvement (Faroqi-Shah et al., 2010) and cross-language improvement in L2 (Croft et al., 2011; Galvez & Hinckley, 2003; Gil & Goral, 2004) relative to those receiving therapy in their L2.
- 2) To what extent do individual-level factors predict naming outcomes in the treated and untreated language? For these analyses, we focused on the effects of baseline naming severity adjusted for participant age on change in naming accuracy over time. We declined to investigate (i) education given difficulties in isolating it from other measures (e.g., cognitive reserve) and (ii) time post-onset given that all participants in our study were in the chronic phase of recovery. We

hypothesized that age and baseline naming severity would modulate treatment outcomes in both languages, such that BWA who demonstrated milder naming difficulties and were younger would show larger gains at the end of treatment.

- 3) To what extent do stimulus-level factors predict naming outcomes in the treated language? For these analyses, we focused on lexical frequency, phonological length, and phonological neighborhood density because the effects of these factors have already been studied in relation to word retrieval in monolingual aphasia (Castro et al., 2020; Gordon, 2002; Gordon & Dell, 2001; Kittredge et al., 2008; Laganaro et al., 2006; Middleton & Schwartz, 2011; Nickels & Howard, 1995, 2004; Nozari et al., 2010) but have yet to be considered in bilingual aphasia. Additionally, these factors were readily extracted from Spanish and English psycholinguistic databases for the treatment stimuli in our dataset whereas other psycholinguistic variables were available only in one language or only for a subset of items. In line with previous studies, we expected more improvement over time for treated words which had: (i) higher lexical frequency, (ii) shorter phonological length, and (iii) denser phonological neighborhood values relative to those with lower frequency, longer phonological length, and sparser phonological neighborhood values.

2. Methods

2.1. Study design

This study examined language treatment data collected as part of the ongoing PROCoM randomized controlled trial (RCT) registered at www.ClinicalTrials.gov (identifier: NCT02916524) described in detail elsewhere (Peñaloza et al., 2020b). The PROCoM RCT aims to evaluate the efficacy of the BiLex computational model (Peñaloza et al., 2019) in predicting treatment outcomes in BWA in the treated and the untreated language and comparing these predicted outcomes to identify the optimal language for rehabilitation in a large sample including 48 Spanish–English BWA (Peñaloza et al., 2020b). Specifically, in this RCT, participants are randomly assigned to a model-prescribed (experimental) group to receive therapy in the language identified as ‘optimal’ in accordance with the model’s prediction (e.g., Spanish) or to a model-non-prescribed (control) group in which therapy is administered opposite the model’s recommendation (e.g., English). Regardless of group assignment, all study participants complete identical cognitive-linguistic assessment batteries pre and post treatment in each language and receive the same intervention (i.e., SFT) for word retrieval deficits across 10 weeks.

The present study examined naming treatment outcomes (irrespective of group assignment) for a subset of these participants within the larger RCT. Naming accuracy for treated and untreated items was assessed separately using naming probes in both languages across 3 pre-treatment (baseline), 10 treatment, and 3 post-treatment sessions for a total of 16 sessions. The uniform and longitudinal nature of this data collection permitted us to evaluate the influence of various intervention-, individual-, and stimulus-level factors on naming accuracy at key time points during treatment. Given that the trial was ongoing at the time of reporting this study, all analyses were conducted on the data of 34 BWA who had completed study participation, and the authors remained blinded to their status in the RCT (i.e., randomization and allocation of participants to the experimental or the control

group). Hence, only within-group results with the known language trained (i.e., Spanish or English) are presented here.

2.2. Participants

Thirty-four Spanish–English bilingual individuals with chronic post-stroke aphasia (i.e., at least six months post-stroke) who completed SFT for word retrieval deficits through the PROCoM RCT were enrolled in this study. Sixteen of these participants have been reported on previously for a different study examining treatment delivery format (Peñaloza et al., 2021). The mean age of the participants was 52.07 years ($SD = 16.47$, $range = 18.7-82.4$) and their mean number of years of education was 13.82 ($SD = 2.95$, $range = 7-19$). The majority of them ($n = 28$) reported Spanish as their L1 and their mean L2 age of acquisition was 11.85 years ($SD = 9.17$, $range = 0-35$). Participants were recruited from a variety of sites across the U.S. including Massachusetts ($n = 13$), California ($n = 9$), Texas ($n = 7$), Washington ($n = 1$), Rhode Island ($n = 1$), Connecticut ($n = 1$), North Carolina ($n = 1$), and from Canada ($n = 1$) as well. Recruitment was managed through Boston University and two secondary sites at San Francisco State University in San Francisco, CA and Austin Speech Labs in Austin, TX. Each participant worked with Spanish–English bilingual clinicians and staff to complete study procedures either in person at one of the recruitment sites or at home via videoconference (see Peñaloza et al., 2021 for a detailed demonstration of comparable study procedures, treatment effectiveness, and reliability across the two therapy delivery modalities). None of the participants presented with pre-morbid neurological disorders or active medical conditions which would have affected their ability to complete treatment. Additionally, all had normal or corrected-to-normal vision and hearing and demonstrated adequate comprehension to support testing and treatment. Informed written consent was obtained from each participant undergoing study procedures and was reviewed and approved by the Boston University Charles River Campus Institutional Review Board (reference number: 4492E). The demographic and clinical backgrounds of the participants are reported in Table 1.

Assessment of language impairment

Participants completed a variety of cognitive-linguistic assessments pre- and post-treatment (see Peñaloza et al., 2020b for the full list of standardized language tests); we highlight the most relevant to the current study in this section. The Western Aphasia Battery-Revised (WAB-R; Kertesz, 2006) and its Spanish version (Kertesz & Pascual-Leone García, 1990) were used to establish the presence of aphasia and to compute the aphasia quotient (WAB-AQ) as an index of severity in each language. The Boston Naming Test in English (BNT; Kaplan et al., 2001) and Spanish (Kohnert et al., 1998) were used to quantify word retrieval impairment via confrontation naming in each language. In addition, the Pyramids and Palm Trees (PAPT; Howard & Patterson, 1992) was administered in the participant’s preferred language to assess semantic knowledge. In general, testing sessions occurred twice per week. Each session lasted about two hours and was blocked by language to minimize cross-language interference which may have arisen from language-switching between tasks. The language of administration was counterbalanced across sessions. Individual pre-treatment WAB-AQ, BNT, and PAPT scores are presented in Table 1.

Table 1. Demographic and clinical characteristics of the Spanish-English BWA.

ID	Sex	Treated Language	Age (years)	Education	Months Post Onset	PAPT	L1		L2	
							WAB-AQ	BNT	WAB-AQ	BNT
P1	M	Spn	82.4	16	401.1	42.3%	55.7	13.3%	29.6	1.7%
P2	F	Spn	54.6	17	58.8	86.5%	74.1	36.7%	68.5	38.3%
P3	M	Spn	44.5	16	19.5	90.4%	84.5	46.7%	89.8	78.3%
P4	F	Eng	24.9	16	6.3	80.8%	27.3	1.7%	37.3	28.3%
P5	F	Spn	26.6	12	129.6	88.5%	77.5	35.0%	67.6	41.7%
P6	M	Spn	68.3	16	244.7	92.3%	82.0	51.7%	67.6	40.0%
P7	F	Spn	47.2	19	53.2	94.2%	79.1	63.3%	54.4	13.3%
P8	M	Spn	53.5	17	37.9	88.5%	51.3	21.7%	47.5	6.7%
P9	M	Spn	77.2	16	26.8	96.2%	67.4	51.7%	64.7	50.0%
P10	F	Eng	78.5	11	38.5	92.3%	78.9	45.0%	76.8	40.0%
P11	M	Spn	70.5	12	6.0	92.3%	57.3	36.7%	39.8	8.3%
P12	F	Spn	27.4	14	48.6	80.8%	72.3	23.3%	66.4	15.0%
P13	F	Spn	53.9	16	44.4	98.1%	90.0	90.0%	68.8	40.0%
P14	M	Eng	69.3	12	10.3	92.3%	35.9	10.0%	46.5	18.3%
P15	F	Eng	55.6	19	41.0	94.2%	96.5	90.0%	60.8	13.3%
P16	M	Eng	62.7	10	23.9	69.2%	9.5	0.0%	11.4	0.0%
P17	F	Spn	47.4	9	12.6	86.5%	82.4	53.3%	71.2	38.3%
P18	M	Eng	56.7	9	51.5	92.3%	81.2	31.7%	91.0	80.0%
P19	M	Eng	39.6	13	40.3	88.5%	21.0	6.7%	39.5	5.0%
P20	M	Eng	42.7	12	22.7	90.4%	94.6	68.3%	57.8	16.7%
P21	M	Spn	62.7	16	52.9	96.2%	78.6	38.3%	85.2	83.3%
P22	F	Eng	21.5	12	23.4	88.5%	34.4	6.7%	53.3	31.7%
P23	M	Eng	63.7	14	385.1	73.1%	47.7	36.7%	15.4	1.7%
P24	M	Eng	55.7	12	46.6	94.2%	97.2	85.0%	92.6	55.0%
P25	F	Eng	67.8	15	8.3	46.2%	21.7	0.0%	23.9	0.0%
P26	F	Spn	50.6	17	31.7	98.1%	92.6	53.3%	69.8	33.3%
P27	F	Eng	44.5	17	34.6	92.3%	85	68.3%	36.5	8.3%
P28	M	Eng	53.5	12	6.8	80.8%	41.4	5.0%	40.1	23.3%
P29	M	Eng	58.8	7	8.9	57.7%	14.7	0.0%	7.8	0.0%
P30	F	Eng	38.4	16	7.0	88.5%	68.3	43.3%	82.6	61.7%
P31	M	Spn	33.0	10	9.9	82.7%	76.1	51.7%	20.5	1.7%
P32	F	Eng	57.0	14	69.9	88.5%	60.5	15.0%	89.6	80.0%
P33	M	Eng	62.5	13	9.5	94.2%	85.4	70.0%	61.7	36.7%
P34	M	Eng	18.7	13	13.9	92.3%	27.0	1.7%	72.6	35.0%
Mean			52.1	13.8	59.6	85.6%	63.2	36.8%	56.1	30.1%
SD			16.5	3.0	95.4	13.4%	26.1	27.4%	24.0	25.4%

Demographic information, pre-treatment language impairment scores, and summary statistics are reported for the 34 study participants. *Age* = age at time of consent; *PAPT* = Percentage score on the Pyramids and Palm Trees test (max score = 52); *L1* = first acquired language; *L2* = second acquired language; *WAB-AQ* = aphasia quotient from the Western Aphasia Battery-Revised (max. score = 100); *BNT* = Percentage score on the Boston Naming Test (max. score = 60).

Assessment of pre-stroke language proficiency

Participants also completed the Language Use Questionnaire (LUQ; Kastenbaum et al., 2019) to characterize their pre-stroke language proficiency in both English and Spanish. The LUQ

was administered in the participant's preferred language and caregivers and/or family members were encouraged to supplement an individual's responses when necessary. The LUQ probes bilingual background across a variety of constructs that contribute to

proficiency in each language including (i) L2 AGE OF ACQUISITION which reflected the participants' age of second language learning onset; (ii) DAILY USAGE computed as the average amount of time (in hours) a participant spent utilizing each language during weekdays and weekends; (iii) FAMILY PROFICIENCY, calculated as the average confidence ratings for each parent and siblings in both languages; (iv) EDUCATIONAL HISTORY, measured as the amount of time the participant used and was exposed to each language across levels of schooling; (v) LIFETIME EXPOSURE computed as the average percentage of time spent speaking, listening, and reading in each language; (vi) LIFETIME CONFIDENCE or the percentage of confidence developed in speaking, listening, and reading in each language; and (vii) LANGUAGE ABILITY RATINGS reflecting self-reported measures of skill in each language. In this study, Spanish and English LUQ metrics (available in Table 2) were coded in L1 and L2 reflecting the first and second acquired languages as done in our previous work (Carpenter et al., 2020; Peñaloza et al., 2019). LUQ metrics for a portion of these participants are also reported in Marte et al., 2022).

2.3. Stimuli

Items

Treatment items were selected for each participant on the basis of incorrect naming attempts in BOTH English and Spanish during pre-treatment administration of a large naming screener of 273 pictured items organized into 13 broad semantic categories with validated semantic features (Sandberg et al., 2020). Cognates and words with at least 50% phonetic similarity were excluded from testing and selection. Additionally, clinicians collected familiarity ratings for 'no-response' items to avoid selecting words which were not named due to differences in word learning or language exposure rather than post-stroke anomia. Therefore, the chosen stimuli for each individual reflected a subset of words from each participant's premorbid vocabulary in their two languages that were inconsistently available or inaccessible after ABI. Stimuli were organized into six sets (15 items in each): set 1 included trained items (e.g., apple), set 2, semantically-related items (e.g., orange), and set 3, unrelated control items (e.g., horse), in the treated language (e.g., English) as well as their corresponding translation sets (*manzana-naranja-caballo* respectively) in the untreated language (e.g., Spanish). For this study, only naming responses for trained and control items (sets 1 and 3) were analyzed in the treated and untreated languages.

Naming probes

Direct effects on treated items and cross-language generalization effects on untreated translations were assessed through a series of 16 picture-naming probes: 3 during pre-treatment testing, 10 during the treatment phase, and 3 during post-treatment testing after treatment was withdrawn. Naming probes were always administered at the beginning of each session and the order of pictures was randomized across sessions and constrained such that (i) no items from the same semantic category appeared next to one another and (ii) no more than two items from the trained sets appeared next to one another. All items in English and Spanish were probed in the same session, but assessment was blocked by language and the order of administration was counterbalanced across sessions. Once all pictures were assessed in one language, the bilingual clinician engaged the participant in a short conversation in the upcoming target language to minimize possible cross-language interference effects associated with blocked

language testing and then presented the pictures in the next language. Naming probes were scored at the item level as correct or incorrect according to previously reported criteria (Kiran et al., 2014).

Psycholinguistic variables

To examine the effects of stimulus-level psycholinguistic properties on treatment outcomes, lexical frequency, phonological length, and phonological neighborhood density values were extracted for the trained stimuli. Lexical frequency was expressed as the standardized occurrence of a word per one million words in the language. Phonological length was reported as the number of constituent phonemes in a word and phonological neighborhood density was measured as the number of words which differed from a target word by a single phoneme addition, deletion, or substitution. In English, these values were gathered from the Cross-Linguistic Easy Access Resource for Phonological and Orthographic Neighborhood Densities (CLEARPOND; Marian et al., 2012) database. Equivalent values in Spanish were collected from CLEARPOND and BuscaPalabras, or B-Pal, a database program with a default vocabulary of 31,491 words (Davis & Perea, 2005). Items were excluded from the stimulus-level analyses if psycholinguistic data was not available in the databases, resulting in 223 items with defined values for the analyses conducted here (see Appendix S1 for the full list of stimuli).

2.4. Treatment

Participants received 40 hours (2 hours per treatment session, twice per week for 10 weeks) of computerized, SFT for word retrieval deficits in one language (i.e., English or Spanish) adapted from our previous work in bilingual aphasia (Edmonds & Kiran, 2006; Kiran et al., 2013; Kiran & Iakupova, 2011; Kiran & Roberts, 2010). As described earlier, the treatment language was decided by random assignment according to the aims pursued in the RCT (see section 2.1). During therapy, participants completed a series of treatment steps for each trained item aiming to facilitate retrieval. Treatment steps and requirements were as follows: (i) *object naming*: naming of the trained item picture; (ii) *feature selection*: identification of semantic features that apply or do not apply to the trained item out of a list of 15 features and classification of features that apply according to type (function, general characteristics, physical attributes, location, and superordinate category); (iii) *association*: generation of an association between the trained item and another item or concept and elaboration on how the two are related; (iv) *yes/no questions*: review of 15 semantic features and decision about whether they apply or do not apply to the trained item; (v) *naming*: naming of the trained item picture; (vi) *sentence production*: generation of a short sentence including the trained item. Full details about the technical requirements and setup for treatment, clinical procedures for administration, and modifications to support remote therapy delivery are available elsewhere (Peñaloza et al., 2021; Scimeca et al., 2022).

2.5. Data Analysis

Data were analyzed using a series of generalized linear mixed-effects models (GLMMs) to estimate the likelihood of a correct response to items in the naming probes. GLMMs were selected because they work well with binomially distributed outcomes (e.g., accuracy data) and are able to account for sources

Table 2. Language background characteristics for the Spanish-English BWA.

ID	L1						L2						
	Use	Fam	Edu	Exp	Con	LAR	AoA	Use	Fam	Edu	Exp	Con	LAR
P1	0.62	1.00	0.89	0.80	1.00	1.00	35.00	0.38	0.17	0.11	0.20	0.36	0.80
P2	0.55	1.00	0.89	0.67	1.00	1.00	6.00	0.45	0.67	0.11	0.33	0.28	0.54
P3	0.01	1.00	0.50	0.26	0.65	0.66	6.00	0.99	1.00	0.50	0.74	0.95	0.86
P4	0.14	0.75	0.28	0.47	0.66	0.89	5.00	0.86	0.58	0.72	0.53	0.68	1.00
P5	0.91	1.00	0.39	0.53	0.72	0.80	5.00	0.09	0.75	0.61	0.47	0.62	0.80
P6	0.29	1.00	0.83	0.62	1.00	0.89	27.00	0.71	0.25	0.17	0.38	0.47	0.91
P7	0.09	1.00	1.00	0.60	1.00	1.00	18.00	0.91	0.17	0.00	0.40	0.45	0.91
P8	0.02	1.00	1.00	0.68	1.00	1.00	12.00	0.98	0.08	0.00	0.32	0.46	1.00
P9	0.24	1.00	0.83	0.43	1.00	0.97	18.00	0.76	0.42	0.17	0.57	0.58	1.00
P10	0.45	1.00	0.58	0.55	1.00	1.00	10.00	0.55	0.25	0.42	0.45	0.56	0.97
P11	0.70	1.00	0.75	0.81	1.00	1.00	15.00	0.30	0.00	0.25	0.19	0.48	0.89
P12	0.37	1.00	0.56	0.63	0.97	1.00	11.00	0.63	0.75	0.44	0.37	0.47	1.00
P13	0.74	1.00	0.94	0.59	1.00	1.00	3.00	0.26	0.42	0.06	0.41	0.51	0.74
P14	0.25	0.92	0.00	0.50	0.81	1.00	3.00	0.75	0.83	1.00	0.50	0.98	1.00
P15	1.00	1.00	0.83	0.90	1.00	1.00	13.00	0.00	0.00	0.17	0.10	0.16	0.31
P16	0.66	0.92	0.83	0.67	1.00	1.00	16.00	0.34	0.13	0.17	0.33	0.42	0.83
P17	0.60	1.00	0.92	0.69	1.00	1.00	16.00	0.40	0.42	0.08	0.31	0.48	0.80
P18	0.75	1.00	0.75	0.31	0.91	0.86	5.00	0.25	0.42	0.25	0.69	0.93	0.74
P19	0.35	1.00	0.50	0.76	1.00	1.00	21.00	0.65	0.17	0.50	0.24	0.44	1.00
P20	0.81	0.92	0.06	0.85	0.88	1.00	5.00	0.19	1.00	0.94	0.15	1.00	1.00
P21	0.26	1.00	0.17	0.33	1.00	0.80	8.00	0.74	0.75	0.83	0.67	0.92	1.00
P22	0.30	0.92	0.33	0.39	0.59	1.00	3.00	0.70	0.42	0.67	0.61	0.60	1.00
P23	0.76	1.00	1.00	0.84	1.00	1.00	3.00	0.24	1.00	0.00	0.16	0.67	0.49
P24	0.71	0.92	0.78	0.90	0.96	0.80	5.00	0.29	0.83	0.22	0.10	0.74	0.77
P25	0.77	1.00	1.00	0.83	1.00	1.00	21.00	0.23	0.42	0.00	0.17	0.49	0.93
P26	0.47	1.00	1.00	0.74	0.90	1.00	16.00	0.53	0.63	0.00	0.26	0.47	0.77
P27	0.93	1.00	0.94	0.68	1.00	1.00	0.00	0.07	0.75	0.06	0.32	1.00	0.83
P28	0.05	1.00	0.25	0.57	1.00	1.00	9.00	0.95	0.50	0.75	0.43	0.89	1.00
P29	0.96	1.00	1.00	0.84	1.00	1.00	17.00	0.04	0.33	0.00	0.16	0.50	0.83
P30	0.26	1.00	0.00	0.55	1.00	1.00	8.00	0.74	0.42	1.00	0.45	0.78	1.00
P31	0.82	1.00	1.00	0.98	1.00	1.00	24.00	0.18	0.00	0.00	0.02	0.05	0.37
P32	0.14	1.00	0.28	0.15	0.43	0.60	1.00	0.86	0.67	0.72	0.85	0.96	1.00
P33	0.87	1.00	1.00	0.98	1.00	1.00	35.00	0.13	0.33	0.00	0.02	0.13	0.63
P34	0.36	0.75	0.00	0.29	0.69	1.00	3.00	0.64	0.83	1.00	0.71	0.91	1.00
Mean	0.51	0.97	0.65	0.63	0.92	0.95	11.85	0.49	0.48	0.35	0.37	0.60	0.84
SD	0.30	0.06	0.35	0.21	0.15	0.10	9.17	0.30	0.30	0.35	0.21	0.26	0.19

Note. All language background metrics reported were measured by the Language Use Questionnaire (Kastembaum et al., 2019). Data for some of these participants are also reported in Marté et al., 2022. Spanish was reported as L1 for all participants except for P13, P15, P20, P23, P24, and P27 (L1=English). Proficiency and history metrics are expressed as the proportion of time spent using a given language (Use, Edu, Exp) or as self-ratings of ability (Fam, LAR) or confidence (Con) in the language.

L1 = first acquired language; L2 = second acquired language; Use = pre-stroke daily usage; Fam = family proficiency; Edu = educational history and usage; Exp = lifetime exposure; Con = lifetime confidence; LAR = language ability rating; AoA = age of acquisition.

of variation in the data unrelated to the experimental design (Jaeger, 2008). Furthermore, these methods are frequently used to model grouped data structures such as repeated observations

nested within participant and item (Baayen et al., 2008; Gordon, 2019; Quicke et al., 2019) which is particularly relevant for the treatment data in this study. The GLMMs were

constructed with a logit link function using the lme4 package (Bates et al., 2015) in R Version 4.0.2 (R Core Team, 2020). Results from the GLMMs were extracted, converted from log-odds to probabilities, and plotted using the following R packages: tidyverse (Wickham et al., 2019), broom (Robinson & Hayes, 2020), patchwork (Pedersen, 2020), and effects (Fox, 2003; Fox & Weisberg, 2018).

Model structure

In each model, item-level accuracy on the naming probes (i.e., scored as 0 or 1) served as the dependent measure. The random effects structure included: (i) random intercepts for participant and item to allow for differences in pre-treatment naming accuracy according to individual participant and item characteristics and (ii) a by-participant random slope for 'session' to capture varying rates of improvement among the participants as a function of their time in the intervention (Gilmore et al., 2022). Other random slopes were excluded from the analyses either because our hypotheses about the treatment data did not support their inclusion or because convergence issues prevented us from overtly specifying them (Barr et al., 2013). Fixed effects for the intervention-level models in aim 1 were composed of (i) session (i.e., continuous probe number with 0-2 for pre-treatment, 3-12 for treatment, and 13-15 for post-treatment), (ii) item set (i.e., trained vs. control items in the treated language and trained translations vs. control translations in the untreated language), (iii) treatment language (i.e., L1 or L2), and (iv) a three-way interaction between these predictors. Individual-level models in aim 2 included fixed effects of (i) session, (ii) treatment language, (iii) initial naming severity (i.e., pre-treatment z-score on the BNT), (iv) a three-way interaction between these predictors, and (v) a covariate for age (i.e., at time of consent expressed in years, centered at the mean). To examine whether outcomes for the trained items in the treated language were influenced by stimulus-level properties, models in aim 3 incorporated fixed effects of (i) session, (ii) one of three psycholinguistic variables (i.e., continuous predictors for lexical frequency, phonological length, and phonological neighborhood density), and (iii) a two-way interaction between these predictors. After visualizing the distributions of the three psycholinguistic variables and before fitting the models, log-transformations were applied to lexical frequency and phonological neighborhood density to mitigate the potential effects of large variable scales on the regression output (see Table S1 and Figure S1 for summary statistics). Finally, a stimulus-language variable was added to control for natural psycholinguistic differences between English and Spanish (e.g., on-average longer word length in Spanish) and their possible effects on naming accuracy. Across all models, categorical predictors were contrast-coded to facilitate the interpretation of the coefficients and discuss results as main effects in the data. Model structures and variable coding procedures are summarized in Table 3 and Table S2.

3. Results

3.1. Intervention- and individual-level factors that modulate language treatment outcomes in BWA

The first series of models examined the effects of intervention-level (i.e., treatment language and number of treatment sessions completed) and individual-level (i.e., baseline naming severity and age) factors on naming outcomes in the treated and untreated

languages. Models were fitted separately, and results were grouped and presented below according to outcomes in the treated and untreated languages.

Outcomes in the treated language

In the intervention-level model, there was a significant interaction between session and item set (collapsing across treatment language), indicating higher likelihood of an accurate response for the trained items over the course of treatment ($b = 0.13$, $SE = 0.01$, $p < .001$). Additionally, this treatment effect was modulated by treatment language; participants who received therapy in their L1 demonstrated greater session-by-session outcomes than did those who completed treatment in their L2 ($b = 0.02$, $SE = 0.01$, $p < .001$). The significance of this three-way session \times item set \times treatment language interaction further suggests that the L1 benefit emerged in concert with the intervention, selectively boosting the naming accuracy of trained words throughout the duration of treatment. Intervention-level factors and their effects on treated language outcomes are illustrated below in Figure 1.

In the next model, we assessed the influence of individual-level factors on naming outcomes via a three-way interaction between session, treatment language, and naming severity (i.e., pre-treatment BNT scores) and a covariate for age, focusing only on the trained items. At baseline, there was a significant effect of naming severity ($b = 0.87$, $SE = 0.30$, $p < .01$), such that individuals with higher BNT scores demonstrated higher likelihood of a correct response. This relationship also persisted over time as participants with higher naming performance experienced more treatment gains than did individuals with poorer scores ($b = 0.28$, $SE = 0.06$, $p < .001$). However, there was no evidence of a three-way interaction including treatment language ($b = -0.14$, $SE = 0.09$, $p = .12$). One possible interpretation of this result is that the relationship between naming severity and time exerted the same effect on treatment outcomes regardless of treatment language. Finally, we did not detect a main effect of age on naming accuracy ($b = -0.02$, $SE = 0.01$, $p = .06$) which together with the observed treatment effects, suggests that the intervention was effective for participants regardless of their age at treatment onset. Naming severity and its effect on predicted accuracy for the trained items over time is shown in Figure 2.

Outcomes in the untreated language

The intervention-level model in the untreated language revealed a significant interaction between session and item set, suggesting an increase in the likelihood of a correct response for the untreated translations of the trained items over time ($b = 0.04$, $SE = .01$, $p < .001$). As expected, this generalization effect was weaker, resulting in modest naming improvement rather than the robust outcomes associated with the treatment effect in the treated language models. Figure 3 demonstrates a slight trend toward greater accuracy for items when L2 was the untreated language which suggests that individuals who experienced greater treatment gains following therapy in L1 also showed larger generalization effects. However, this pattern was not observed over time, indicating instead that generalization outcomes were mostly similar across L1 and L2 following a non-significant interaction between session, item set, and treatment language (or in this case, untreated language; $b = 0.01$, $SE = 0.01$, $p = .09$).

Given that treatment language was not an informative factor in the intervention-level model, we first determined whether to retain the term in the individual-level model. The results of a likelihood ratio test for model comparison suggested treatment

Table 3. Proposed regression analyses with fixed and random effects coding.

Model	Outcome	Fixed Effects	Fixed Effects Coding	Random Effects
1a: Treatment-level (Treated Language)	Item accuracy for Trained and Control Items	Session	Continuous	(Session Participant) (1 Item)
		Item Set	Contrast Coding (-1 for Control Items, 1 for Trained Items)	
		Treatment Language	Contrast Coding (-1 for L2, 1 for L1)	
1b: Treatment-level (Untreated Language)	Item accuracy for Translations of Trained and Control Items	Session	Continuous	(Session Participant) (1 Item)
		Item Set	Contrast Coding (-1 for Control Items, 1 for Trained Items)	
		Treatment Language	Contrast Coding (-1 for L2, 1 for L1)	
2a: Individual-level (Treated Language)	Item accuracy for Trained Items	Session	Continuous	(Session Participant) (1 Item)
		Treatment Language	Contrast Coding (-1 for L2, 1 for L1)	
		Naming Severity	Continuous (z-scored BNT)	
		Age	Covariate, Continuous (centered)	
2b ¹ : Individual-level (Untreated Language)	Item accuracy for Translations of Trained Items	Session	Continuous	(Session Participant) (1 Item)
		Treatment Language	Contrast Coding (-1 for L2, 1 for L1)	
		Naming Severity	Continuous (z-scored BNT)	
		Age	Covariate, Continuous (centered)	
3a: Lexical Frequency	Item accuracy for Trained Items	Session	Continuous	(Session Participant) (1 Item)
		Log Frequency	Continuous, Log-transform	
		Stimulus Language	Covariate, Contrast Coding (-1 for English, 1 for Spanish)	
3b: Phonological Length	Item accuracy for Trained Items	Session	Continuous	(Session Participant) (1 Item)
		Length	Continuous	
		Stimulus Language	Covariate, Contrast Coding (-1 for English, 1 for Spanish)	
3c: Phonological Neighborhood Density	Item accuracy for Trained Items	Session	Continuous	(Session Participant) (1 Item)
		Log Density	Continuous, Log-transform	
		Stimulus Language	Covariate, Contrast Coding (-1 for English, 1 for Spanish)	

Note. ¹The model structure for 2b was altered slightly given the outcomes from model 1b. Justification is provided in the Results section. *Log frequency* = log-transform of lexical frequency; *Log density* = log-transform of phonological neighborhood density.

language was unlikely to improve overall fit ($\chi^2(4) = 4.12, p = .39$), and therefore we excluded this factor from the final individual-level analysis in the untreated language.

At baseline in the individual-level model, there was a significant effect of naming severity ($b = 1.73, SE = 0.29, p < .01$) such that milder aphasia severity was associated with greater likelihood of a correct naming response in the untreated language. As in the treated language, the effect of naming severity also interacted with time, suggesting that larger generalization effects emerged for individuals with higher baseline naming performance than for participants who scored worse on the BNT ($b = 0.11, SE = 0.03, p < .001$). Finally, there was no significant effect of age on untreated language outcomes ($b = 0.01, SE = 0.01, p = .26$), suggesting that generalization effects were equally likely to occur for participants regardless of age.

3.2. Stimulus-level factors that modulate language treatment outcomes in BWA

The final series of models examined the influence of the psycholinguistic variables on naming accuracy in the treated language. Given that both English and Spanish treated data were included in these models, we conducted independent samples t-tests to determine if there were any between-language differences in the distributions of the psycholinguistic measures before completing the regression analyses. Results revealed significant differences in the samples for log frequency, phonological length, and phonological neighborhood density across the two languages (see Figure S1). Therefore, we included stimulus language (i.e., Spanish or English) as a predictor in each of the psycholinguistic models presented below to account for this known variation in word properties.

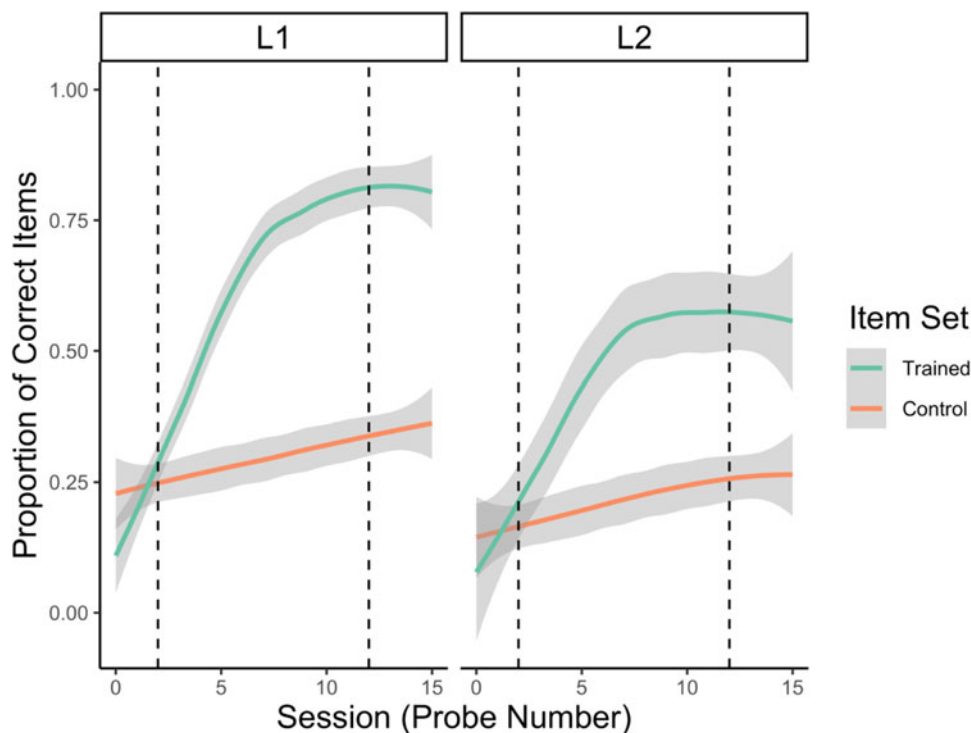


Figure 1. Overall change in naming accuracy in the treated language over the course of treatment. Average proportion of correct items named by the BWA in the treated language (L1 versus L2) across 16 naming probes is depicted over the course of treatment. Naming accuracy is shown separately for trained items (green) and control items (orange). Gray shading represents the standard error of the model prediction. Session denotes naming probes during the baseline phase (sessions 0-2), the treatment phase (sessions 3-12), and the post-treatment phase (sessions 13-15).

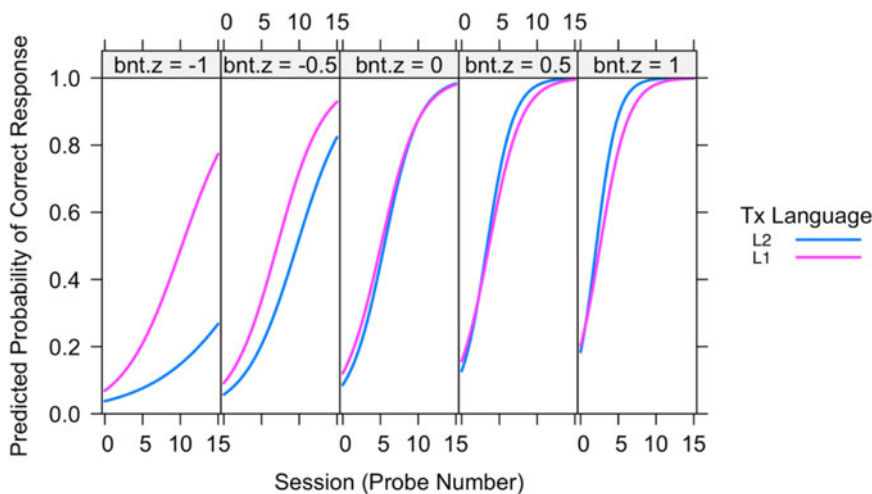


Figure 2. Effect of baseline naming severity on change in naming accuracy for trained items in the treated language over the course of treatment. The effect of baseline naming impairment on the predicted probability of a correct naming response for trained items in the treated language is depicted over the course of treatment. Effects are shown for L1 (pink) and L2 (blue) separately at various degrees of initial naming severity (z-score values on the BNT: -1, -0.5, 0, 0.5, and 1). Lower initial severity values correspond to larger treatment effects and this effect does not differ between L1 and L2. *Tx Language* = Treatment Language.

The first psycholinguistic model revealed a significant effect of log frequency at baseline ($b = 0.92$, $SE = 0.22$, $p < .001$) and no main effect of stimulus language ($b = 0.01$, $SE = 0.28$, $p = .95$). These findings indicate that words with higher frequency values were more likely to be named correctly before treatment

regardless of language. Additionally, log frequency did not contribute to change in naming accuracy over time which suggests that the effect of this lexical-semantic property on word retrieval did not persist as participants completed more treatment sessions ($b = 0.004$, $SE = 0.01$, $p = .79$).

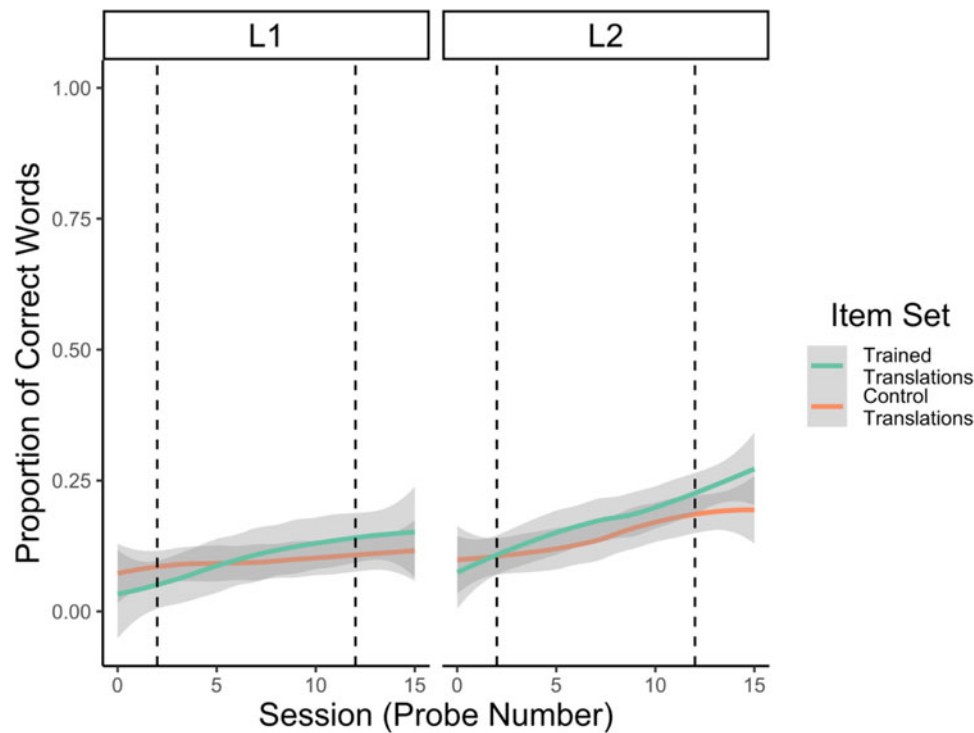


Figure 3. Overall change in naming accuracy in the untreated language over the course of treatment. Average proportion of correct words produced by the BWA in the untreated language (L1 versus L2) across 16 naming probes over the course of treatment. Naming accuracy is shown separately for translations of trained items (green) and translations of control items (orange). Gray shading represents the standard error of the model prediction. Session denotes naming probes during the baseline phase (sessions 0-2), the treatment phase (sessions 3-12), and the post-treatment phase (sessions 13-15).

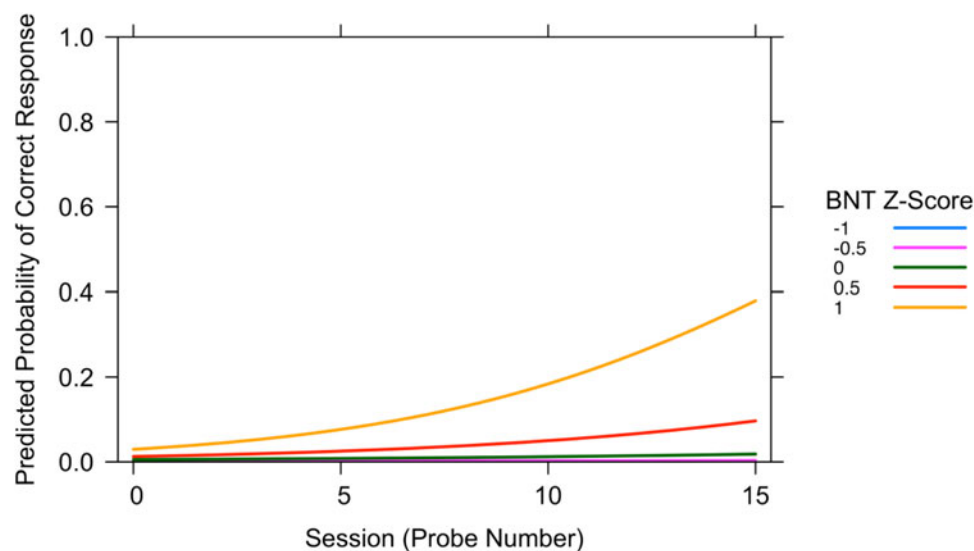


Figure 4. Effect of baseline naming severity on change in naming accuracy for translations of the trained items in the untreated language over the course of treatment.

The effect of baseline naming impairment on the predicted probability of a correct naming response for translations of the trained items in the untreated language (collapsed across L1 and L2) is depicted over the course of treatment. Colored lines represent predictions across different degrees of initial naming severity (BNT z-scores ranging between -1 and 1). Lower naming severity suggests greater generalization to translations of the trained items in the untreated language over time.

In contrast to the log frequency model, the phonological length model revealed no effect of word length at baseline across the two languages ($b = 0.06$, $SE = 0.08$, $p = .39$), suggesting that at pre-treatment, words were equally difficult to name regardless of their length. However, a significant length effect did emerge over time, indicating

lower expected accuracy for longer words when compared to shorter words ($b = -0.03$, $SE = 0.01$, $p < .001$) during treatment. Finally, there was no main effect of stimulus language ($b = -0.16$, $SE = 0.30$, $p = .57$) which again suggests that word retrieval accuracy was unaffected by length differences in Spanish and English.

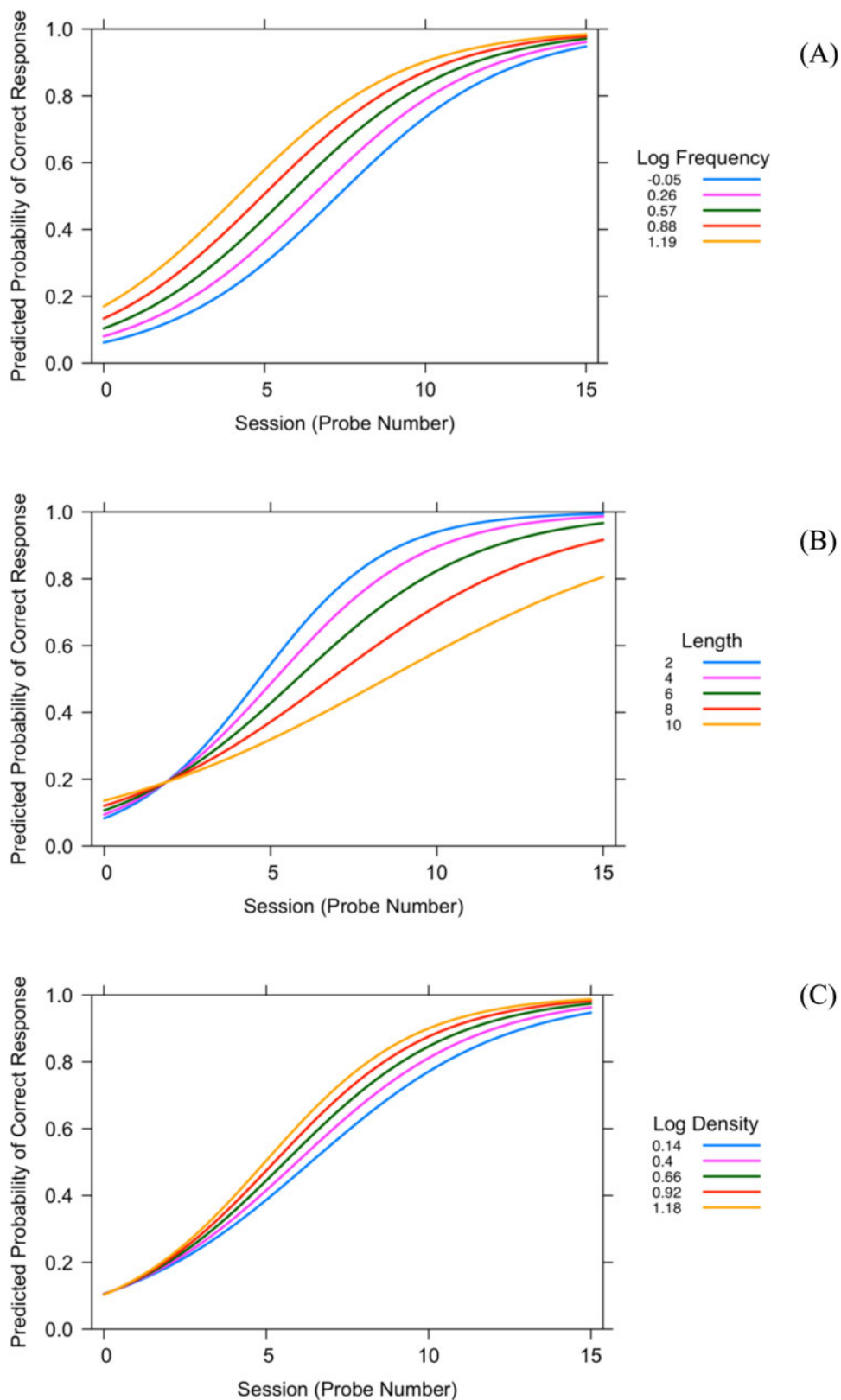


Figure 5. Effects of frequency, phonological length, and phonological neighborhood density on change in naming accuracy for trained items in the treated language over the course of treatment.

The effects of the three psycholinguistic features on the predicted probability of a correct response for trained items in the treated language are depicted across three panels (collapsed across English and Spanish). Panel (A) depicts the effects of lexical frequency, with Log Frequency values plotted for the mean as well as ± 0.5 standard deviations and ± 1 standard deviations. This figure shows that frequency predicts naming responses at baseline but not over time. Panel (B) shows the effects of phonological length, with length values plotted for representative word length values. This figure indicates that length does not predict accuracy at baseline, but that shorter words are more accurately named over treatment. Panel (C) depicts the effects of phonological neighborhood density, with log density values plotted for the mean as well as ± 0.5 standard deviations and ± 1 standard deviations. This figure suggests that phonological neighborhood density does not predict accuracy at baseline, but words with denser neighborhoods are more accurately named over time.

The results from the log neighborhood density model were largely in line with findings from the phonological length model. At baseline, there was no significant effect of log neighborhood density across the two languages ($b = -0.02$, $SE = 0.28$, $p = .92$), suggesting that naming accuracy before treatment did not differ across words with respect to the size of their phonological neighborhoods. Nevertheless, a session by log neighborhood density interaction demonstrated that over time, treatment words with denser phonological neighborhoods were more likely to be named correctly ($b = 0.09$, $SE = 0.02$, $p < .001$). Once again, there was no main effect of stimulus language ($b = -0.08$, $SE = 0.29$, $p = .77$) which further indicates that the relationship between word retrieval accuracy and phonological neighborhood density was not affected by differences in Spanish and English.

4. Discussion

The present study aimed to examine intervention-, individual-, and stimulus-level treatment parameters and their relative contributions to language therapy outcomes for a group of 34 Spanish–English BWA who received SFT for word retrieval deficits. In the sections which follow, we discuss in more detail the effects of these multilevel factors on session-by-session changes in naming accuracy over the course of treatment across participants and items and highlight the implications of our findings for treatment design and delivery in populations with bilingual aphasia.

In general, our results revealed that the SFT provided in one language was effective at improving word retrieval deficits for Spanish–English BWA across the treated and the untreated languages. In the treated language, participants demonstrated large increases in naming accuracy for trained items relative to untrained control items over the course of intervention. A similar yet smaller magnitude of improvement emerged in the untreated language over time in which accuracy for the untreated translations of the trained items was higher when compared to accuracy for the untreated translations of the control items. These results expand the existing evidence of the effectiveness of SFT for aphasia (see Efstratiadou et al., 2018; Quique et al., 2019 for a review) and extend the evidence for treatment effectiveness of lexical-semantic therapy approaches for within-language improvement and cross-language generalization in bilingual aphasia (Croft et al., 2011; Edmonds & Kiran, 2006; Kiran et al., 2013; Kiran & Iakupova, 2011; Kiran & Roberts, 2010; Peñaloza et al., 2021). To our knowledge, this study is the first in bilingual aphasia rehabilitation to examine the magnitude of incremental changes in word retrieval abilities for nouns (see Li et al., 2020 for similar verb-retrieval results) via careful analysis of data over the course of treatment at the individual item-level. Given that GLMMs were used in our analyses, the results may also be considered generalizable across participants and items due to the nature of the random effects structures specified in the models (Baayen et al., 2008; Gordon, 2019; Quique et al., 2019). Future studies which deliver SFT to Spanish–English BWA in a manner consistent with our procedures should expect similar outcomes based on our measurements of treatment effects across participants and items.

The observed patterns of improvement in both languages were further explained by intervention-level factors including number of treatment sessions and treatment language. In the treated language, we found that the likelihood of a correct naming response for a trained item was higher over time when therapy was provided in L1. More importantly, this L1 effect during treatment was highly specific, as it only emerged for the trained items but

not for the untrained control items. This finding suggests that the observed gains in naming accuracy among the participants resulted from the SFT intervention rather than repeated item exposure during the naming probes (i.e., repeated testing carry-over effects). The observation that larger treatment outcomes emerged when treatment was provided in L1 aligns with previous studies which have also pointed to an L1 advantage following various treatment approaches for BWA (Croft et al., 2011; Kiran et al., 2013; also see Faroqi-Shah et al., 2010 for a review). From a theoretical view, increased direct treatment effects for L1 therapy are also supported by bilingual models of lexical processing and word retrieval. For example, the Revised Hierarchical Model (Kroll & Stewart, 1994) posits that word retrieval in L1 is facilitated by activation flowing from the conceptual-semantic system to the L1 mental lexicon via robust associative connections shaped by a strong L1 proficiency. Although word retrieval in L2 may also be semantically-mediated in cases of high L2 proficiency, the connections between the semantic-conceptual system and the L2 lexicon are expected to be weaker than those with the L1 lexicon. During SFT, practice involving the semantic features of trained words is assumed to lead to increased activation of their representations in the semantic-conceptual system which may directly enhance the activation of the target words at the lexical and phonological level, making them more readily available for retrieval (Quique et al., 2019). Thus, increased accessibility of trained words in the treated L1 via strengthened semantics may ultimately reflect better recovery for L1 at the specific item-level as a result of treatment.

In the untreated language, there was no significant effect of treatment language suggesting that, when they occur, cross-language generalization outcomes do not depend on the language targeted in therapy. We did, however, observe an increasing – albeit non-significant – trend in the accuracy of L2 trained translations over time. This suggests that BWA who demonstrated an L1 advantage in the treated language were also likely to show more generalization to their untreated L2. Integrating this result into the current evidence on cross-language generalization effects after language therapy for bilingual aphasia is challenging given the mixed previous findings. On one hand, some studies have signaled greater generalization to L2 when L1 was targeted in therapy (Croft et al., 2011; Galvez & Hinckley, 2003; Gil & Goral, 2004). On the other, some prior studies have observed greater generalization effects in L1 following therapy in L2 (Edmonds & Kiran, 2006; Kiran & Iakupova, 2011). Although the data presented here do not directly endorse either group of findings, they pattern more closely with the first group of studies which pointed to more generalization to L2 when L1 was the treatment language. Nevertheless, the most appropriate conclusion is that in the absence of a clear advantage, both treatment in L1 and L2 resulted in statistically significant cross-language generalization to the other language. Overall, these findings may have important clinical implications for treatment planning and delivery. On one hand, if an individual's goal is to maximize word retrieval in one specific language, our results suggest that more improvement in the treated language would be expected when that language was L1. If, however, treatment goals are to improve outcomes across both languages, an individual could be counseled to pursue therapy in L1 or L2 as treatment in either language might be expected to result in comparable amounts of cross-language generalization to the other based on our results.

Baseline naming severity was also an important predictor of naming accuracy across the treated and untreated languages,

confirming the notion that the prognostic value of initial language impairment on language recovery in monolinguals with aphasia (see Kristinsson et al., 2022; Quique et al., 2019 for a review) also extends to BWA. In the treated language, higher pre-treatment BNT scores indicating milder anomia were associated with larger improvement on trained items over time. However, there was no effect of treatment language on this interaction which suggested that the effect of naming severity was comparable across L1 and L2. Given that both English and Spanish treatment data were included in the same models, it should be noted that the effect of baseline naming impairment on change in naming accuracy during treatment may be interpreted independent of the language in which the assessment was administered. Stated differently, lower initial naming severity on the English BNT and the Spanish BNT both predicted larger treatment outcomes at the individual-level. These severity effects are supported by a recent systematic review of semantic-feature analysis studies in monolingual aphasia which reported similar baseline anomia effects as measured by the BNT on naming outcomes (Quique et al., 2019).

In the untreated language, lower baseline scores on the BNT were also associated with higher likelihood of a correct naming response for untreated translations of the trained items over time. This study is one of the first to demonstrate the predictive value of a clinical assessment for word retrieval on item-level cross-language generalization outcomes following language therapy for BWA. Altogether, these results suggest that the BNT is an effective indicator of eventual treatment response for Spanish–English BWA regardless of which language they receive therapy in.

Of note, age did not affect naming accuracy in the treated nor the untreated language. This ultimately suggests that direct treatment improvement and cross-language generalization were equally likely to occur for BWA regardless of their age at treatment enrollment. Although age has been found to modulate aphasia recovery (Lazar et al., 2010; Plowman et al., 2012; Watila & Balarabe, 2015; see also Kristinsson et al., 2022) there is evidence that age does not modulate the effects of semantic-based treatments for word retrieval deficits in monolinguals with aphasia (Quique et al., 2019). Our study supports the latter evidence and suggests that treatment-induced recovery in both treated and untreated language occurs independently from age in BWA.

Rehabilitation outcomes were also influenced by the properties of the treatment words themselves. In the treated language, two distinct patterns of psycholinguistic effects on naming accuracy emerged: a lexical-semantic effect (i.e., frequency) on baseline naming accuracy and a phonological effect (i.e., phonological length and neighborhood density) on naming over time. It should be noted that at baseline, words in the treatment set were difficult to name across participants as they were specifically selected on the basis of two incorrect naming trials (once in Spanish and once in English). Therefore, the effect of log frequency at baseline suggests that a small number of words in the individual treatment sets may have been spontaneously named correctly because their high log frequency values conferred lower thresholds for activation, thereby facilitating production. By contrast, the expected facilitatory effects of lower phonological length (Castro et al., 2020; Nickels & Howard, 1995, 2004) and greater neighborhood density (Gordon, 2002; Gordon & Dell, 2001; Middleton & Schwartz, 2011) on word retrieval were not observed on baseline naming which suggests that initial word retrieval accuracy before treatment was not modulated by phonological structure. However, these effects were observed over time, as lower phonological length and greater neighborhood density did predict better

outcomes for the trained items. Two possible explanations are available for this observed dissociation between the lexical-semantic effect at baseline and the phonological effects over time. First, the manner in which treatment effects are typically measured in naming studies heavily relies on phonological output processes. Second, given that the underlying cause of anomia may be different across individuals with aphasia (Howard & Gatehouse, 2006; Lambon-Ralph et al., 2002), it may be the case that for the majority of the BWA in our study, the locus of their naming impairment was at the level of accessing phonological representations stored in the phonological output lexicon (i.e., post-semantic). During SFT, semantic word representations are enhanced by the intervention tasks (e.g., feature selection) which should be indexed by lexical-semantic properties such as lexical frequency. However, if anomic deficits occur more post-semantically, semantic activation – strengthened via therapy – may propagate to the correct lexical entries but fail to activate the correct phonological nodes to facilitate successful single word production (Dell et al., 1997). This is possibly why the phonological factors are more important and overtake the effects of frequency as predictors of treatment effects OVER TIME. These post-semantic naming deficits might be more visible in future analyses which could consider treatment effects in naming not as binarized values of correct versus incorrect but instead as improvements in the number of phonemes produced correctly in trained items over time.

4.1 Limitations and a path for future research

Some limitations of this work should be considered. Although this was a prospective study with a pre-defined set of predictor variables, language treatment effects are multifactorial, and other factors may also play an important role in determining therapy response in both languages in BWA. Indeed, a variety of other factors could be initialized at the various levels outlined in our study to determine their effects such as more fine-grained measures of time spent in therapy (e.g., minutes of therapy adjusted for breaks or number of items completed per session) or other pre-treatment assessment scores for semantic and/or phonological loci of impairment. Additionally, while we carefully measured language proficiency in the two languages with the LUQ at pre-treatment, language use and history factors were not directly accounted for in these analyses. Understanding the potential relationships between these proficiency variables and treatment effects represents an important avenue of future research which could improve treatment planning and assist us in accounting for individual differences in treatment outcomes.

For the psycholinguistic models, we were forced to drop treatment words from the analyses for which data could not be retrieved from linguistic databases. In the future, aphasia rehabilitation studies may benefit from more widely available lexical databases which include relevant features for items typically used in word retrieval studies. These issues aside, future studies may continue this work to deepen our understanding of how psycholinguistic variables may affect word retrieval and treatment outcomes in diverse ways. For example, one additional line of research would be to examine errors in lexical retrieval before, during, and after treatment alongside stimulus-level properties to examine whether there is any relation between the quantity and quality of errors at certain psycholinguistic values. Another useful follow-up would be to investigate treatment effects with

additional psycholinguistic predictors such as familiarity, imageability, and naming agreement.

5. Conclusions

In sum, our study upholds the effectiveness of SFT for BWA and demonstrates that multilevel factors such as number of treatment sessions, treatment language, pre-treatment naming severity, and psycholinguistic values may be useful in predicting rehabilitation outcomes. Our findings demonstrate that Spanish–English BWA may show word retrieval improvement over time in both L1 and L2 with varying degrees of post-stroke severity and across diverse treatment stimuli. These results also have implications for clinical practice and treatment planning, especially for selecting the target language in therapy and constructing sets of treatment stimuli. Clinicians may look to these findings when developing personalized rehabilitation plans for bilingual clients and when interpreting treatment outcomes.

Acknowledgements. This research was supported by U01 DC014922 awarded to Swathi Kiran. Additionally, Michael Scimeca was partially supported by T32 DC013017 and Claudia Peñaloza was supported by the Juan de la Cierva-Incorporación Program (IJC2018-037818) funded by Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación MCIN/AEI 10.13039/501100011033. The authors wish to thank the participants and their families for study participation and the members of the Boston University Aphasia Research Lab for their time spent collecting and analyzing data.

Data availability statement. The data and analyses that support the findings of this study are available from the authors upon request.

Competing interests. Swathi Kiran serves as a cofounder and advisor to Constant Therapy Health. There is no scientific overlap with this work and that entity.

Supplementary material. For supplementary material accompanying this paper, visit <https://doi.org/10.1017/S1366728923000391>

References

- Abutalebi, J., Rosa, P., Tettamanti, M., Green, D., & Cappa, S. (2009). Bilingual aphasia and language control: A follow-up fMRI and intrinsic connectivity study. *Brain and Language*, 109, 141–156. <https://doi.org/10.1016/j.bandl.2009.03.003>
- Ansaldo, A. I., & Saidi, L. G. (2014). Aphasia therapy in the age of globalization: Cross-linguistic therapy effects in bilingual aphasia. *Behavioural Neurology*, 2014, 603085. <https://doi.org/10.1155/2014/603085>
- 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>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 10.1016/j.jml.2012.11.001. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Carpenter, E., Rao, L., Peñaloza, C., & Kiran, S. (2020). Verbal fluency as a measure of lexical access and cognitive control in bilingual persons with aphasia. *Aphasiology*, 34, 1–22. <https://doi.org/10.1080/02687038.2020.1759774>
- Castro, N., Stella, M., & Siew, C. S. Q. (2020). Quantifying the interplay of semantics and phonology during failures of word retrieval by people with aphasia using a multiplex lexical network. *Cognitive Science*, 44(9), e12881. <https://doi.org/10.1111/cogs.12881>
- Centeno, J. (2008). Multidisciplinary evidence to treat bilingual individuals with aphasia. *Perspectives on Communication Disorders and Sciences in Culturally and Linguistically Diverse Populations*, 15, 66–72. <https://doi.org/10.1044/cds15.3.66>
- Centeno, J. (2009). Issues and principles in service delivery to communicatively impaired minority bilingual adults in neurorehabilitation. *Seminars in Speech and Language*, 30, 139–152. <https://doi.org/10.1055/s-0029-1225951>
- Costa, A., La Heij, W., & Navarrete, E. (2006). The dynamics of bilingual lexical access. *Bilingualism: Language and Cognition*, 9(2), 137–151. <https://doi.org/10.1017/S1366728906002495>
- Croft, S., Marshall, J., Pring, T., & Hardwick, M. (2011). Therapy for naming difficulties in bilingual aphasia: Which language benefits? *International Journal of Language & Communication Disorders*, 46(1), 48–62. <https://doi.org/10.3109/13682822.2010.484845>
- Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, 37(4), 665–671. <https://doi.org/10.3758/BF03192738>
- 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, 801–838.
- Edmonds, L. A., & Kiran, S. (2006). Effect of semantic naming treatment on crosslinguistic generalization in bilingual aphasia. *Journal of Speech, Language, and Hearing Research*, 49(4), 729–748. [https://doi.org/10.1044/1092-4388\(2006/053\)](https://doi.org/10.1044/1092-4388(2006/053))
- Efstratiadou, E. A., Papatasiou, I., Holland, R., Archonti, A., & Hilari, K. (2018). A systematic review of semantic feature analysis therapy studies for aphasia. *Journal of Speech, Language, and Hearing Research*, 61(5), 1261–1278.
- Fabbro, F. (2001). The bilingual brain: Bilingual aphasia. *Brain and Language*, 79(2), 201–210. <https://doi.org/10.1006/brln.2001.2480>
- Faroqi-Shah, Y., Frymark, T., Mullen, R., & Wang, B. (2010). Effect of treatment for bilingual individuals with aphasia: A systematic review of the evidence. *Journal of Neurolinguistics*, 23(4), 319–341. <https://doi.org/10.1016/j.jneuroling.2010.01.002>
- Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, 8, 1–27. <https://doi.org/10.18637/jss.v008.i15>
- Fox, J., & Weisberg, S. (2018). Visualizing fit and lack of fit in complex regression models with predictor effect plots and partial residuals. *Journal of Statistical Software*, 87, 1–27. <https://doi.org/10.18637/jss.v087.i09>
- Galvez, A., & Hinckley, J. J. (2003). Transfer patterns of naming treatment in a case of bilingual aphasia. *Brain and Language*, 87(1), 173–174. [https://doi.org/10.1016/S0093-934X\(03\)00256-6](https://doi.org/10.1016/S0093-934X(03)00256-6)
- Gil, M., & Goral, M. (2004). Nonparallel recovery in bilingual aphasia: Effects of language choice, language proficiency, and treatment. *International Journal of Bilingualism*, 8(2), 191–219. <https://doi.org/10.1177/13670069040080020501>
- Gilmore, N., Mirman, D., & Kiran, S. (2022). Young adults with acquired brain injury show longitudinal improvements in cognition after intensive cognitive rehabilitation. *Journal of Speech, Language, and Hearing Research*, 65(4), 1494–1520. https://doi.org/10.1044/2021_JSLHR-21-00324
- Goral, M., & Lerman, A. (2020). Variables and mechanisms affecting response to language treatment in multilingual people with aphasia. *Behavioral Sciences*, 10(9), 144.
- Gordon, J. (2002). Phonological neighborhood effects in aphasic speech errors: Spontaneous and structured contexts. *Brain and Language*, 82(2), 113–145. [https://doi.org/10.1016/S0093-934X\(02\)00001-9](https://doi.org/10.1016/S0093-934X(02)00001-9)
- Gordon, K. (2019). How mixed-effects modeling can advance our understanding of learning and memory and improve clinical and educational practice. *Journal of Speech, Language, and Hearing Research*, 62(3), 507–524. https://doi.org/10.1044/2018_JSLHR-L-ASTM-18-0240
- Gordon, J., & Dell, G. S. (2001). Phonological neighborhood effects: Evidence from aphasia and connectionist modeling. *Brain and Language*, 79(1), 21–23. <https://doi.org/10.1006/brln.2001.2574>
- Grasemann, U., Peñaloza, C., Dekhtyar, M., Miikkulainen, R., & Kiran, S. (2021). Predicting language treatment response in bilingual aphasia using neural network-based patient models. *Scientific Reports*, 11, 10497. <https://doi.org/10.1038/s41598-021-89443-6>

- Hendricks, C. T., Nicholas, M. L., & Zipse, L. (2014). Effects of phonological neighbourhood on the treatment of naming in aphasia. *Aphasiology*, 28(3), 338–358. <https://doi.org/10.1080/02687038.2013.866209>
- Hinckley, J. J. (2003). Picture naming treatment in aphasia yields greater improvement in L1. *Brain and Language*, 87(1), 171–172. [https://doi.org/10.1016/S0093-934X\(03\)00255-4](https://doi.org/10.1016/S0093-934X(03)00255-4)
- Howard, D., & Gatehouse, C. (2006). Distinguishing semantic and lexical word retrieval deficits in people with aphasia. *Aphasiology*, 20(9), 921–950. <https://doi.org/10.1080/02687030600782679>
- Howard, D., & Patterson, K. (1992). *Pyramids and Palm Trees Test*. Pearson.
- Hughes, E., Roberts, J., & Tainturier, M.-J. (2012). Cross-linguistic generalisation of treatment in Welsh-English bilingual anomia. *Procedia - Social and Behavioral Sciences*, 61, 168–169. <https://doi.org/10.1016/j.sbspro.2012.10.131>
- Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, 59(4), 434–446. <https://doi.org/10.1016/j.jml.2007.11.007>
- Junqué, C., Vendrell, P., Vendrell-Brucet, J. M., & Tobeña, A. (1989). Differential recovery in naming in bilingual aphasics. *Brain and Language*, 36(1), 16–22. [https://doi.org/10.1016/0093-934X\(89\)90049-7](https://doi.org/10.1016/0093-934X(89)90049-7)
- Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *The Boston Naming Test*. Pro-Ed.
- 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>
- Keane, C., & Kiran, S. (2015). The nature of facilitation and interference in the multilingual language system: Insights from treatment in a case of trilingual aphasia. *Cognitive Neuropsychology*, 32(3–4), 169–194. <https://doi.org/10.1080/02643294.2015.1061982>
- Kertesz, A. (2006). *Western Aphasia Battery-Revised*. PsychCorp.
- Kertesz, A., & Pascual-Leone García, A. (1990). *Batería de afasias "Western" [Western Aphasia Battery en versión y adaptación castellana]*. Nau Llibres.
- 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>
- Kiran, S., Sandberg, C., Gray, T., Ascenso, E., & Kester, E. (2013). Rehabilitation in bilingual aphasia: Evidence for within and between-language generalization. *American Journal of Speech-Language Pathology*, 22(2), S298–S309. [https://doi.org/10.1044/1058-0360\(2013\)12-0085](https://doi.org/10.1044/1058-0360(2013)12-0085)
- Kiran, S., Balachandran, I., & Lucas, J. (2014). The nature of lexical-semantic access in bilingual aphasia. *Behavioural Neurology*, 2014, 389565. <https://doi.org/10.1155/2014/389565>
- Kittredge, A., Dell, G., Verkuilen, J., & Schwartz, M. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, 25(4), 463–492. <https://doi.org/10.1080/02643290701674851>
- 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>
- Kohnert, K. (2009). Cross-language generalization following treatment in bilingual speakers with aphasia: A review. *Seminars in Speech and Language*, 30(3), 174–186. <https://doi.org/10.1055/s-0029-1225954>
- Kohnert, K. J., Hernandez, A. E., & Bates, E. (1998). Bilingual performance on the Boston naming test: Preliminary norms in Spanish and English. *Brain and Language*, 65(3), 422–440. <https://doi.org/10.1006/brln.1998.2001>
- Kristinsson, S., den Ouden, D. B., Rorden, C., Newman-Norlund, R., Neils-Strunjas, J., & Fridriksson, J. (2022). Predictors of therapy response in chronic aphasia: Building a foundation for personalized aphasia therapy. *Journal of Stroke*, 24(2), 189–206. <https://doi.org/10.5853/jos.2022.01102>
- Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connection between bilingual memory representations. *Journal of Memory and Language*, 33(2), 149–174. <https://doi.org/10.1006/jmla.1994.1008>
- 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, 445. <https://doi.org/10.3389/fpsyg.2019.00445>
- Laganaro, M., Di Pietro, M., & Schnider, A. (2003). Computerised treatment of anomia in chronic and acute aphasia: An exploratory study. *Aphasiology*, 17(8), 709–721. <https://doi.org/10.1080/02687030344000193>
- Laganaro, M., Di Pietro, M., & Schnider, A. (2006). What does recovery from anomia tell us about the underlying impairment: The case of similar anomic patterns and different recovery. *Neuropsychologia*, 44(4), 534–545. <https://doi.org/10.1016/j.neuropsychologia.2005.07.005>
- Laine, M., & Martin, N. (2006). *Anomia: Theoretical and clinical aspects*. Psychology Press.
- Lambon-Ralph, M. A., Moriarty, L., & Sage, K. (2002). Anomia is simply a reflection of semantic and phonological impairments: Evidence from a case-series study. *Aphasiology*, 16(1–2), 56–82.
- Lazar, R. M., Minzer, B., Antoniello, D., Festa, J. R., Krakauer, J. W., & Marshall, R. S. (2010). Improvement in aphasia scores after stroke is well predicted by initial severity. *Stroke*, 41(7), 1485–1488. <https://doi.org/10.1161/STROKEAHA.109.577338>
- Li, R., Li, W., & Kiran, S. (2020). Effect of Mandarin Verb Network Strengthening Treatment (VNeST) in Mandarin-English bilinguals with aphasia: A single-case experimental design. *Neuropsychological Rehabilitation*, 31(8), 1224–1253. <https://doi.org/10.1080/09602011.2020.1773278>
- Lorenzen, B., & Murray, L. L. (2008). Bilingual aphasia: A theoretical and clinical review. *American Journal of Speech-Language Pathology*, 17(3), 299–317. [https://doi.org/10.1044/1058-0360\(2008\)026](https://doi.org/10.1044/1058-0360(2008)026)
- Marangola, P., Rizzi, C., Peran, P., Piras, F., & Sabatini, U. (2009). Parallel recovery in a bilingual aphasic: A neurolinguistic and fMRI study. *Neuropsychology*, 23(3), 405–409. <https://doi.org/10.1037/a0014824>
- Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLOS ONE*, 7(8), e43230. <https://doi.org/10.1371/journal.pone.0043230>
- Marte, M. J., Carpenter, E., Falconer, I. B., Scimeca, M., Abdollahi, F., Peñalosa, C., & Kiran, S. (2022). LEX-BADAT: Language EXperience in Bilinguals With and Without Aphasia DATaset. *Frontiers in Psychology*, 13, 875928. <https://doi.org/10.3389/fpsyg.2022.875928>
- Meinzer, M., Obleser, J., Flaisch, T., Eulitz, C., & Rockstroh, B. (2007). Recovery from aphasia as a function of language therapy in an early bilingual patient demonstrated by fMRI. *Neuropsychologia*, 45(6), 1247–1256. <https://doi.org/10.1016/j.neuropsychologia.2006.10.003>
- Middleton, E. L., & Schwartz, M. F. (2011). Density pervades: An analysis of phonological neighbourhood density effects in aphasic speakers with different types of naming impairment. *Cognitive Neuropsychology*, 27(5), 401–427. <https://doi.org/10.1080/02643294.2011.570325>
- Miller Amberger, A. (2012). Language intervention in French–English bilingual aphasia: Evidence of limited therapy transfer. *Journal of Neurolinguistics*, 25(6), 588–614. <https://doi.org/10.1016/j.jneuroling.2011.10.002>
- Moss, A., & Nicholas, M. (2006). Language rehabilitation in chronic aphasia and time postonset. *Stroke*, 37(12), 3043–3051. <https://doi.org/10.1161/01.STR.0000249427.74970.15>
- Nakagawa, Y., Sano, Y., Funayama, M., & Kato, M. (2019). Prognostic factors for long-term improvement from stroke-related aphasia with adequate linguistic rehabilitation. *Neurological Sciences*, 40(10), 2141–2146. <https://doi.org/10.1007/s10072-019-03956-7>
- Nardo, D., Holland, R., Leff, A. P., Price, C. J., & Crinion, J. T. (2017). Less is more: Neural mechanisms underlying anomia treatment in chronic aphasic patients. *Brain*, 140(11), 3039–3054. <https://academic.oup.com/brain/article/140/11/3039/4259065?login=true>
- Nickels, L., & Howard, D. (1995). Aphasic naming: What matters? *Neuropsychologia*, 33(10), 1281–1303. [https://doi.org/10.1016/0028-3932\(95\)00102-9](https://doi.org/10.1016/0028-3932(95)00102-9)
- Nickels, L., & Howard, D. (2004). Dissociating effects of number of phonemes, number of syllables, and syllabic complexity on word production in aphasia: Its the number of phonemes that counts. *Cognitive Neuropsychology*, 21(1), 57–78.

- Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of Memory and Language*, 63(4), 541–559. <https://doi.org/10.1016/j.jml.2010.08.001>
- Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. John Benjamins.
- Pedersen, T. L. (2020). *Patchwork: The composer of plots*. <https://CRAN.R-project.org/package=patchwork>
- Peñaloza, C., & Kiran, S. (2019). Recovery and rehabilitation patterns in bilingual and multilingual aphasia. In J. Schwieter & M. Paradis (Eds.), *The Handbook of the Neuroscience of Multilingualism* (pp. 553–571). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119387725.ch27>
- Peñaloza, C., Grasemann, 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>
- Peñaloza, C., Barrett, K., & Kiran, S. (2020a). 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>
- Peñaloza, C., Dekhtyar, M., Scimeca, M., Carpenter, E., Mukadam, N., & Kiran, S. (2020b). 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>
- Peñaloza, C., Scimeca, M., Gaona, A., Carpenter, E., Mukadam, N., Gray, T., Shamapant, S., & Kiran, S. (2021). Telerehabilitation for word retrieval deficits in bilinguals with aphasia: Effectiveness and reliability as compared to in-person language therapy. *Frontiers in Neurology*, 12, 589330. <https://doi.org/10.3389/fneur.2021.589330>
- Plowman, E., Hentz, B., & Ellis, C. (2012). Post-stroke aphasia prognosis: A review of patient-related and stroke-related factors. *Journal of Evaluation in Clinical Practice*, 18(3), 689–694.
- Quique, Y. M., Evans, W. S., & Dickey, M. W. (2019). Acquisition and generalization responses in aphasia naming treatment: A meta-analysis of semantic feature analysis outcomes. *American Journal of Speech-Language Pathology*, 28(1S), 230–246. https://doi.org/10.1044/2018_AJSLP-17-0155
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org>.
- Roberts, J., & Tainturier, M.-J. (2010). Cross-linguistic treatment generalisation in Welsh-English bilingual anomia. *Procedia - Social and Behavioral Sciences*, 6, 262–263. <https://doi.org/10.1016/j.sbspro.2010.08.133>
- Robinson, D., & Hayes, A. (2020). *broom: Convert statistical analysis objects into tidy tibbles*. <https://CRAN.R-project.org/package=broom>
- Rodriguez, C. J., Allison, M., Daviglius, M. L., Isasi, C. R., Keller, C., Leira, E. C., Palaniappan, L., Piña, I. L., Ramirez, S. M., Rodriguez, B., Sims, M., American Heart Association Council on Epidemiology and Prevention, American Heart Association Council on Clinical Cardiology, & American Heart Association Council on Cardiovascular and Stroke Nursing. (2014). Status of cardiovascular disease and stroke in Hispanics/Latinos in the United States: A science advisory from the American Heart Association. *Circulation*, 130(7), 593–625. <https://doi.org/10.1161/CIR.0000000000000071>
- Sandberg, C., Gray, T., & Kiran, S. (2020). Development of a free online interactive naming therapy for bilingual aphasia. *American Journal of Speech-Language Pathology*, 29(1), 20–29. https://doi.org/10.1044/2019_AJSLP-19-0035
- Sandberg, C. W., Zacharewicz, M., & Gray, T. (2021). Bilingual Abstract Semantic Associative Network Training (BABsANT): A Polish-English case study. *Journal of Communication Disorders*, 93, 106143. <https://doi.org/10.1016/j.jcomdis.2021.106143>
- Scimeca, M., Abdollahi, F., Peñaloza, C., & Kiran, S. (2022). Clinical perspectives and strategies for confronting disparities in social determinants of health for Hispanic bilinguals with aphasia. *Journal of Communication Disorders*, 98, 106231. <https://doi.org/10.1016/j.jcomdis.2022.106231>
- Shafto, M. A., & Tyler, L. (2014). Language in the aging brain: The network dynamics of cognitive decline and preservation. *Science*, 346(6209), 583–587. <https://www.science.org/doi/10.1126/science.1254404>
- Watila, M. M., & Balarabe, S. A. (2015). Factors predicting post-stroke aphasia recovery. *Journal of the Neurological Sciences*, 352(1-2), 12–18.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>