Cross-language generalization of language treatment in multilingual people with post-stroke aphasia

Goral, Mira; Norvik, Monica I; Antfolk, Jan; Agrotou, Ioulia; Lehtonen, Minna

Published in:
Brain and Language

DOI:
10.1016/j.bandl.2023.105326

Published: 01/11/2023

Brain & Language 246 (2023) 105326

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Review

Cross-language generalization of language treatment in multilingual people with post-stroke aphasia: A meta-analysis

Mira Goral [a,b,c,*], Monica I. Norvik [c,d], Jan Antfolk [e], Ioulia Agrotou [b], Minna Lehtonen [c,f,g]

[a] Lehman College, The City University of New York, Bronx, NY, USA
[b] The Graduate Center, The City University of New York, New York, NY, USA
[c] MultiLing Center for Multilingualism in Society Across the Lifespan, Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo, Norway
[d] Department of Acquired Brain Injury, Statped, Oslo, Norway
[e] Department of Psychology, Abo Akademi University, Turku, Finland
[f] Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland
[g] Turku Brain and Mind Center, University of Turku, Turku, Finland

A R T I C L E I N F O

Keywords:
Aphasia
Treatment
Intervention
Multilingual
Bilingual
Cross-language generalization
Meta-analysis

A B S T R A C T

Studies on the efficacy of language treatment for multilingual people with post-stroke aphasia and its generalization to untreated languages have produced mixed results. We conducted a systematic review and a meta-analysis to examine within- and cross-language treatment effects and the variables that affect them. We searched PubMed, PsycINFO, CINAHL, and Google Scholar (February 2020; January 2023), identifying 40 studies reporting on 1573 effect sizes from 85 individuals. We synthesized effect sizes for treatment outcomes using a multi-level model to correct for multiple observations from the same individuals. The results showed significant treatment effects, with robust within-language treatment effects and weaker cross-language treatment effects. Age of language acquisition of the treatment language predicted within-language and cross-language effects. Our results suggest that treating multilingual people with aphasia in one language may generalize to their other languages, especially following treatment in an early-acquired language and a later-learned language that became the language of immersion.

1. Introduction

A question that has long preoccupied the research and clinical aphasia community is whether administering language intervention in one language of multilingual[1] people with aphasia leads to benefits in both the treated and untreated languages. Early as well as more recent research studies have yielded mixed results, leading scholars to formulate refined versions of this question, such as under what circumstances cross-language generalization is found and which variables influence it (e.g., Ansaldo & Ghazi Saidi, 2014; Faroqi-Shah et al., 2010; Kiran et al., 2013; Paradis, 2004). The majority of studies published on this topic present data from single individuals or from small groups of participants, as is typical of intervention studies of the heterogeneous population in question, which adds to the challenge of drawing clear conclusions from a mixed set of data. The purpose of the current systematic review and meta-analysis was to pull together data published to date concerning the question of within- and between-language treatment effects in multilingual people with post-stroke aphasia, with the aim of determining whether intervention is beneficial in all languages and what variables contribute to its efficacy.

Aphasia is an acquired language disorder resulting from insult to the brain. People who acquire aphasia experience varying degrees of difficulty producing and comprehending spoken and written language; difficulty which affects their communication, social engagement, and quality of life (e.g., Hallowell & Chapey, 2008; Papathanasiou et al., 2021). Aphasia severity refers to the degree of impairment, with difficulties ranging on a continuum from mild aphasia, characterized by relatively minimal deficits, to severe aphasia, referring to substantial difficulty expressing oneself and understanding others (e.g., Papathanasiou & Coppens, 2021). In monolingual people, any difficulties measured following the aphasia onset are assumed to be the result of the acquired lesion. People who know multiple languages and sustain brain

* Corresponding author at: Lehman College, 250 Bedford Park Blvd., Bronx, NY 10468, USA.
E-mail address: mira.goral@lehman.cuny.edu (M. Goral).

[1] In this paper we use the adjective *multilingual* to refer to people who use more than one language.

https://doi.org/10.1016/j.bandl.2023.105326
Received 8 October 2022; Received in revised form 9 September 2023; Accepted 26 September 2023
Available online 31 October 2023

lesion resulting in aphasia typically experience language impairments in all their languages. Variation in the degree of impairment and of recovery over time in the languages of multilingual people with aphasia has been noted in the literature (e.g., Albert & Obler, 1978; Kuzmina et al., 2019; Paradis, 1983; 2004). The variation in abilities can be the result of the acquired impairment, but could also reflect less than complete mastery of skills prior to the aphasia onset (e.g., Kiran & Gray, 2018; Lerman et al., 2020).

To account for the variability in abilities across the languages of multilingual people with aphasia, a number of theories have been put forward, including the early hypotheses of Ribot (1881) and Pitres (1895) that the first-acquired language (L1) or the most-used language, respectively, will be better spared in aphasia. Paradis (1983; 1993; 2004) further distinguished patterns of parallel and non-parallel impairment and recovery and argued for their relation to linguistic and meta-linguistic representations of earlier- and later-learned languages. In recent years, theories of language control, including activation and inhibition processes, have been formulated to account for recovery patterns in multilingual people with aphasia, both independent of and in relation to language intervention (e.g., Goral & Lerman, 2020; Green, 1998; Kiran et al., 2013; Paradis, 2004).

People with stroke-induced aphasia typically improve their language abilities with time (especially in the first weeks and months post onset) and with intervention (e.g., Cherney & Robey, 2008; Raymer et al., 2008), although many remain impaired and do not recover their abilities to their pre-onset levels. This is true for the one language of monolingual people with aphasia and for the multiple languages of multilingual people with aphasia. Language intervention in aphasia has been shown to be effective in improving performance, although not eliminating the acquired language deficits (e.g., Brady et al., 2016). Intervention often targets underlying linguistic processes that may account for the observed impairment, or can target specific impaired language abilities, structures, or items.

A central problem that has concerned clinical researchers is that of treatment generalization (e.g., Best et al., 2013; Thompson, 1989; Webster et al., 2015). For language intervention to be most effective, not only the components or items targeted during the intervention should be better accessible to the person with aphasia following treatment, but ideally, improvement will generalize to untreated language components and contexts. These include untargeted aspects of the language being treated (e.g., Best et al., 2013; Quique et al., 2019), as well as the language not targeted directly during therapy (e.g., Lerman et al., 2022). That is, for intervention with multilingual people with aphasia, generalization would ideally be observed not only within the treated language but also from the treated language to comparable aspects in the non-treated language(s). Here, researchers have distinguished between generalization to the direct equivalents of the treated items (e.g., cognate words – translation equivalents that share form in addition to meaning – in naming therapy, Edmonds & Kiran, 2006; Kohnert, 2004) and more general improvements in the non-treated language (e.g., Conner et al., 2018; Goral & Lerman, 2020).

Several review papers have addressed the question of the degree to which cross-language generalization is achieved following therapy in multilingual people with aphasia. Kohnert (2009) conducted a systematic review which yielded 12 studies (10 case studies and two group studies). Kohnert divided her discussion of the results into results of the early publication (1900s) and of the later ones (2000s) and demonstrated that the earlier studies found cross-language generalization more consistently than did the later studies. There was variability in the strength and direction of the cross-language generalization reported across all studies; virtually all studies found within-language generalization. Kohnert implicated several variables as possibly contributing to the mixed pattern of results, including aphasia-related, such as aphasia type and severity, language-related, such as the target linguistic modality and linguistic distance between the languages, and multilingualism-related, such as language use and order of language acquisition.

Faroqi-Shah, Frymark, Mullen, and Wang (2010) published a systematic review of evidence available to answer several questions, two most relevant here: whether treatment is effective when administered in the person's non-L1, and whether treatment administered in one language affects also the person's other language(s). In their review, Faroqi-Shah et al. (2010) included 14 studies with data from 45 participants (one group study of 30 participants, one study with three participants, one study with two participants, and the remaining studies reporting on a single participant). Their results demonstrated that all five studies with intervention in the non-L1 reported treatment benefits. Regarding the question of cross-language generalization, results were mixed, with half the studies included in the review reporting cross-language benefits from intervention in L2 on the untreated L1 (3 of 5 from L2 to L1 for receptive language abilities, and 5 of 11 from L2 to L1 for expressive skills). In the other direction, from L1 to non-L1, there were two studies that measured receptive skills, including the one group study, and both showed generalization, plus all four studies reporting on expressive abilities reported positive cross-language generalization in that direction.

Faroqi-Shah and colleagues discussed these results in the context of several participant- and language-related factors that contributed to the variability in the results. They concluded that neither age of L2 acquisition nor language typology seem to account for the results. A similar conclusion can be found in another review article pertaining to cross-language effects in language intervention in multilingual people with aphasia, published a few years after the Faroqi-Shah et al. paper (Ansaldo & Ghazi Saidi, 2014). Ansaldo and Ghazi Saidi (2014) reviewed 15 articles and focused on the question of cross-language generalization. Their findings are consistent with previous results, demonstrating the potential benefits of treatment in one language on the untreated language. In their review, Ansaldo and Ghazi Saidi (2014) focused on the role of critical variables in contributing to cross-language generalization, including the type of language intervention and its relations to the outcomes measured. Even though Ansaldo and Ghazi Saidi (2014) concurred with Faroqi-Shah et al. (2010) that there is little evidence supporting the role of language distance (Faroqi-Shah et al. considered language distance as determined by whether the languages belonged to the same language family or not; Ansaldo & Ghazi Saidi discussed structural similarities among languages), they did note an effect of item similarity. Specifically, Ansaldo and Ghazi Saidi (2014) demonstrated that studies that compared effects for cognates versus non-cognates showed greater cross-language generalization for cognates. Their review also concurs with Faroqi-Shah et al.'s in concluding that evidence to date is insufficient to determine the role of language abilities – pre- and post- the aphasia onset – in facilitating cross-language generalization.

Whether benefits from intervention are related to the relative abilities in the two or more languages of multilingual people with aphasia has been a matter of debate (e.g., Conner et al., 2018; Goral & Lerman, 2020; Knoph et al., 2015). Researchers have proposed that the spreading activation from a language that was learned after the L1 or from a language of lower proficiency is more likely to occur to the stronger L1, consistent with the Revised Hierarchical Model of bilingual lexical representation (Kroll & Stewart, 1994; Kroll et al., 2010). In contrast, the need to inhibit a stronger language while activating a language of lower proficiency has been proposed to predict lack of cross-language generalization from a non-L1 to an L1 (e.g., Lerman et al., 2022). Additional variables that have been shown to affect treatment efficacy in aphasia, with some mixed results, include aphasia severity (e.g., Kurland et al., 2018; Menahemi-Falkov et al., 2021; Plowman et al., 2011; Quique et al., 2019), and intervention-related variables, such as the target of the intervention and the outcome measures used, have also been examined in the efficacy literature (e.g., Menahemi-Falkov et al., 2021).

Almost twenty years later and in light of the newly accumulated intervention studies with multilingual people with aphasia, we took on conducting a new systematic review, this time combined with a meta-analysis. To our knowledge, this is the first meta-analysis performed on studies that report on treatment effects in multilingual people with aphasia. Unlike previous reviews, the present meta-analysis approach allowed us to quantify the magnitude of treatment effects. To use all available data (i.e., several different measurements taken from a single participant) we accounted for the dependency between observations. We also investigated, and corrected for, publication or reporting bias. Our goal was to examine results from intervention studies to answer the following research questions.

I. Is language treatment with multilingual people with aphasia beneficial?

a) Is treatment effective for the treated language?
b) Do treatment effects generalize to the untreated language(s)?

We hypothesized that performance will be higher post-treatment compared to pre-treatment in the treated language and also, albeit to a lesser extent, in the untreated language.

II. What treatment-, multilingualism-, and aphasia-related variables affect treatment efficacy?

a) Do the targeted treatment levels affect treatment efficacy? Does treatment generalize across language levels? Does treatment generalize across items?
b) Is treatment efficacy different in L1 vs. a later learned language? Is treatment efficacy dependent on participants' age of acquisition of the treatment language? Is treatment efficacy dependent on language proficiency of the treatment language or the relative proficiency of the treated and tested languages?
c) Is cross-language treatment generalization related to language distance of the treated and tested languages?
d) Is treatment efficacy related to aphasia severity?

On the basis of previous findings we hypothesized that a) treatment will be more effective for trained than for untrained items and will generalize best to tasks at the same level that was targeted during treatment; b) treatment in an L1 or a language that was acquired early will be more effective than treatment in a later-learned language and, possibly, that treatment in a proficient language will be more efficacious than treatment in a language of low proficiency; c) cross-language generalization will occur regardless of language distance; and d) treatment efficacy may be dependent of aphasia severity.

2. Method

2.1. Literature search

We searched the electronic databases PubMed, PsycINFO, CINAHL, and Google Scholar in February 2020. The search included terms referring to stroke-induced aphasia, bilingualism or multilingualism, and treatment. For the specific search terms, see Table 1. Before the full search, we tested the sensitivity of the search terms by selecting 10 known papers that fulfilled our inclusion criteria. As those were all found in the search, we deemed that the search was sensitive enough. A second search was conducted in May 2020 without any new hits, and a third, updated search in January 2023 (limiting the search for studies published between 5/2020 and 12/2023). All search hits were screened for their title and abstract, and if potential relevance could not be evaluated based on them, the full-text of the article was referred to. For an overview of the whole screening process, see Fig. 1.

2.2. Inclusion and exclusion criteria

We included studies that reported data on treatment of stroke-induced aphasia in bilingual or multilingual adults. We thus excluded data from participants whose language difficulties were caused by some other origin, such as a tumor or a neurodegenerative disorder. We also excluded data on children or adolescents (participants with age lower than 18 years at the time of the stroke) and participants who spoke two dialects rather than languages. We excluded papers that had only studied spontaneous recovery or used treatments not focusing on language, such as interventions focusing solely on brain stimulation or cognitive training. We only included peer-reviewed journal articles and those written in the English language.

Table 1
Search strings used in the literature search.

Database	Search String
PubMed	(aphasi* OR anomia OR stroke* OR vascular OR hemorrhage) AND (multiling* OR biling* OR triling* OR quadriling* OR polyglot) AND (recover* OR treat* OR intervention OR rehabilitation OR therap*)
PsycINFO	(aphasi* OR anomia OR stroke* OR vascular OR hemorrhage) AND (multiling* OR biling* OR triling* OR quadriling* OR polyglot) AND (recover* OR treat* OR intervention OR rehabilitation OR therap*)
CINAHL	(aphasi* OR anomia OR stroke* OR vascular OR hemorrhage) AND (multiling* OR biling* OR triling* OR quadriling* OR polyglot) AND (recover* OR treat* OR intervention OR rehabilitation OR therap*)
Google Scholar	aphasia OR aphasic OR anomia OR stroke OR vascular OR hemorrhage AND multilingual OR bilingual OR trilingual OR quadrilingual OR polyglot AND recovery OR treatment OR intervention OR rehabilitation OR therapy OR therapeutic

In order for a study to be included, it had to report performance measures from more than one language of the participant, to allow for language generalization effects to be investigated. We included all linguistic measures reported in the papers, but we excluded an overall score of a test where individual subscores were reported separately. Articles without an available full-text (Charlton, 1964) and participants for which no numerical results had been individually reported (e.g., Fredman, 1975; Junqué et al., 1989) were excluded. We only included studies that provided exact values for the measures; we did not extract values from figures. For studies that fulfilled our inclusion criteria but reported no usable data in numerical form, we contacted the authors to obtain more data. We restricted these author inquiries to articles published 20 years ago or later. We got a response about three of the seven studies we inquired about. The final dataset included 89 participants from 40 studies (note also that there were four participants who had participated in two separate studies, amounting to 85 individuals). See Supplementary Table 1.

2.3. Data coding

In the following, we describe the extraction and coding of data from the original articles. See also Supplementary Table 2. We also report procedures for interrater reliability.

2.3.1. Study features, basic demographic, and aphasia characteristics

We coded the following study characteristics: list of authors, year and title of publication, and participant code used in the study. We also screened if the same participants had been studied in different articles: In the final dataset, there were four such participants; they were counted as four, not eight, participants.

We extracted participants' type of aphasia as described in the article, lesion site, and hemisphere. Aphasia severity was determined based on the authors' statements and was coded as an ordinal variable (1 = mild; 2 = mild-to-moderate; 3 = moderate; 4 = moderate-to-severe; 5 = severe). When different severity levels were reported for the two (or more) languages of the participants, we used the severity for the less impaired language. For descriptive purposes, we also extracted participants' age, gender, and education level (1 = lower than high school or upper secondary education, or less than 12 years; 2 = high school or upper secondary education, or 12 years; 3 = university or college education, or more than 12 years).

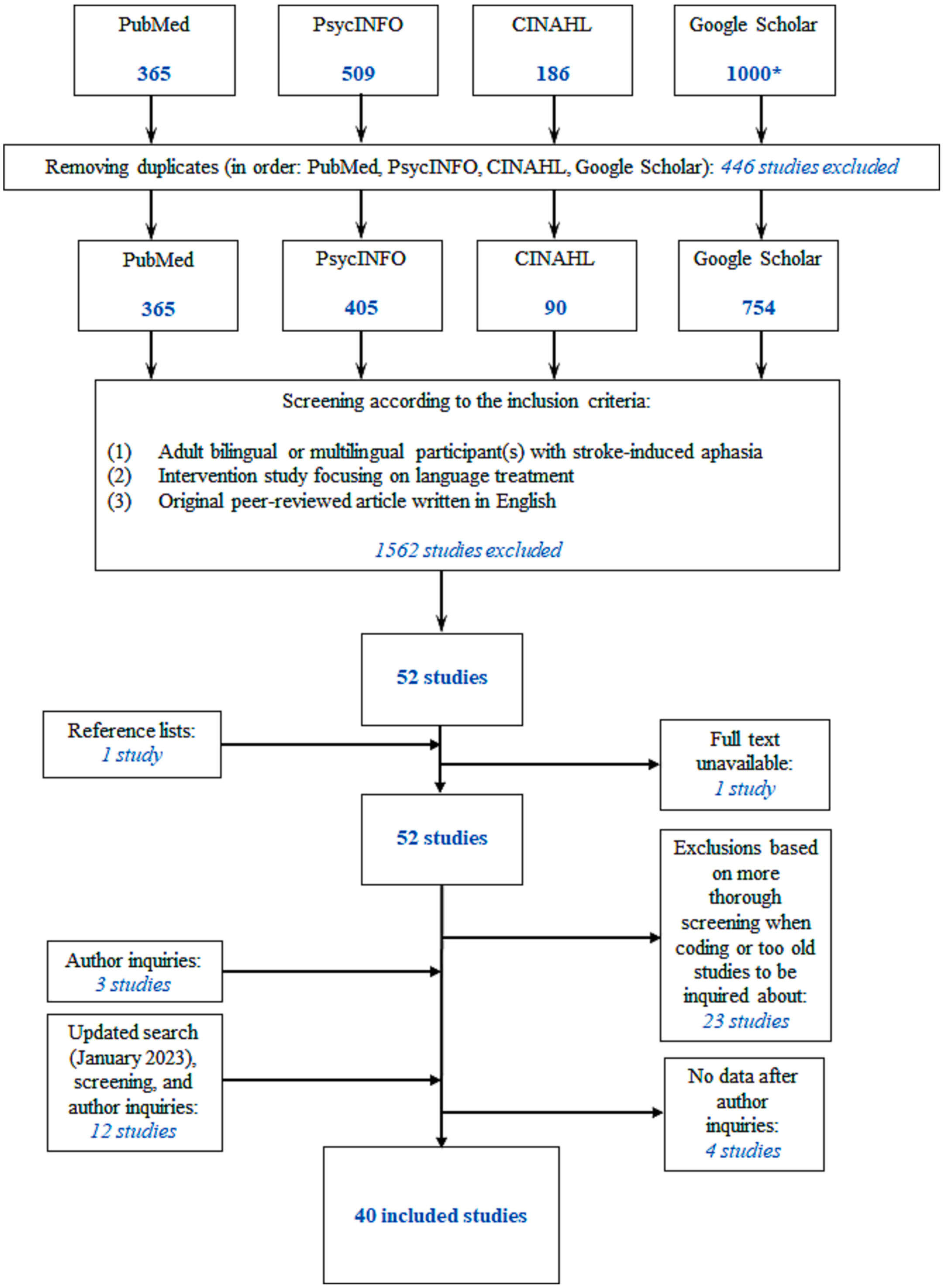

Fig. 1. Flowchart of the screening process.

2.3.2. *Language background*

We extracted information about all the languages of the participants, their acquisition order, age of acquisition, proficiency, and reported use of each language, as described in the articles. Then, we re-coded age of acquisition of each language into four categories (1 = first language(s); 2 = early childhood (under 6 years); 3 = childhood or late childhood (7–18 years); 4 = adulthood (over 18 years)) and proficiency into three categories (1 = low; 2 = medium; 3 = high). Proficiency designation, referring to pre-stroke reported proficiency, was based on the wording reported in the papers: for example, *low, very low* were coded as 1; *medium, mid, good* were coded as 2; *high, very good* were coded as 3. To investigate whether the proficiency difference between the languages affects treatment gains, we also coded the proficiency relationship between the treated and tested language ("tr_higher", when proficiency in the treatment language was higher than that of the test language; "tr_lower", when treatment language proficiency was lower than that of the test language; and "same" when they both belonged to the same proficiency category).

Due to the substantial variation in how language use of the participants was reported in the original articles, this variable could not be reliably categorized for analysis. Finally, to enable studying whether language distance is related to cross-language treatment effects, we coded the participants' languages according to whether they belonged to the same language family and how close they were within a language family (e.g., English is an Indo-European language), that is, how many divisions (e.g., Indo-European, Germanic, Northwest Germanic, West Germanic, North Sea Germanic Anglo-Frisian, Anglic, Later Anglic, Middle Modern English, Macro-English, English) (Hammarström et al., 2022) within a family they share (0–9; 0 = no shared language divisions; 9 = nine shared language divisions). English and Mandarin, for instance, belong to different language families (Indo-European and Sino-Tibetan, respectively), so they were coded as having 0 shared divisions, whereas German and Norwegian belong to the same language family (Indo-European), and also share three subdivisions (Indo-European, Germanic, Northwest Germanic), so they were coded as 3.

2.3.3. Treatment and task characteristics

We coded the language(s) of the treatment and the task, as well as whether the treatment language was (one of) the participants' first language(s) or a later acquired one (L1 vs. L*n*). We also extracted information about the type of treatment as described in the article. There was variability in the interventions used across studies and in the tasks reported for the different interventions. We categorized the treatment type based on the language level they are targeting (i.e., words, sentences, words and sentences, or sentences and discourse). We also coded the tested task domain (production, comprehension, or repetition) and the level of language that it focuses on (subword, word, discourse, or sentence, a combination of these, or an overall score). Then, for each observation, we coded whether the task targets the same language level as the one treatment had targeted, or a different one ("within" or "across" levels). Furthermore, for cases where the task measured cross-level generalization (i.e., "across" levels) we coded whether the task is reflecting a higher level than the treatment (e.g., a sentence-level task for a word-level treatment) or a lower level (e.g., a word-level task for a sentence-level treatment). Here the data allowed four categories: "one step higher" (e.g., from word-level treatment to sentence-level task or from sentence to discourse) or "one step lower" (e.g., from discourse to sentence or sentence to word), "two steps higher" (from word to discourse level) and "two steps lower" (from discourse to word level).

To study the difference between trained vs. untrained items, observations that were testing the same items as those used in treatment were coded as "trained", and the other observations (that reported using different items or for which this was not explicitly stated) were coded as "untrained". For cross-language data, the translation equivalents of the trained items, although not themselves trained, were also coded as "trained" when targeting an item and testing its translation equivalent was explicitly reported by the papers' authors.

Information about frequency and dosage of the intervention was inconsistently reported and we did not include it in our analyses.

2.3.4. Measures

For calculation of effect sizes, we extracted the pre- and post-treatment scores of each task, the maximum score of that task, as well as the number of items used in the task. We also coded these data for follow-up measurements when available. All measures reported by the author(s) of each included paper were used; for example, if the study included values for pre- and post-treatment testing for an authors' experimental task as well as standardized tests, all were included in the analyses.

2.3.5. Interrater reliability

The data of the first search were coded by one coder. Ten randomly selected studies were subjected to calculation of interrater reliability for the variables used in the analyses. These studies were coded independently by two additional coders, and Cohen's Kappa was calculated between the original coding and these coders. For variables that did not reach the preset Cohen's Kappa value of 0.80, two of the authors reviewed all the cells together (and re-coded them when needed), reaching agreement in all of them. For the studies added after the latest search, one author coded the data and a second author reviewed all the data entered; any inconsistencies were resolved.

2.4. Statistical analyses

For statistical analyses, we used the package *metafor* (Viechtbauer, 2010) for *R* (version 4.0.3. for Mac; R Core Team, 2017). The *R* script with the reported analyses, the generated output files, and the data file are available at the project's Open Science Framework domain (https://osf.io/jrh89/).

2.4.1. Calculating effect sizes

Similarly to, for example, Kuzmina et al. (2019), we first calculated the number of correct and incorrect responses in the administered task as a measurement of the relative risk (RR) before and after the intervention. The RR describes the ratio of a probability of an outcome after treatment and the probability of an outcome before treatment. Here, the outcome of interest was the probability of a correct response in the administered task. The probability of a correct response was calculated as the correct responses/total responses. In some instances, outcomes were reported as a percentage and without providing the number of administered items. In these cases, we replaced the total number of items with the median from other tests and used this value to also estimate the absolute number of correct responses based on the reported percentage of correct responses. Before proceeding, we examined and removed rows where the participant's pretest performance was perfect and there was no room to measure improvement. Any row for which there was insufficient information to calculate effect size was excluded. After this, we used the PBIT-argument to estimate the standardized mean difference (SMD) and its standard error. The PBIT-argument calculates the probit transformed risk difference, assuming that the responses are normally distributed on the underlying quantitative scale. The PBIT-argument thus produces a Hedge's *g*, which is a close equivalent of Cohen's *d*, but is less biased when samples are small, and thus sometimes considered a corrected effect size (e.g., Borenstein et al., 2011). We used the *vtype* argument "UB" to obtain unbiased estimates of the sampling variance (Viechtbaue, 2010). In our case, data were coded such that higher positive values describe larger improvement (i.e., better performance after the intervention compared to before the intervention).

2.4.2. Pooling effect sizes within comparisons

In 16 instances, we pooled effect sizes across highly similar outcome measures, assumed to interchangeably reflect the same processes (such as subtests Antonyms I and Antonyms II in the *Bilingual Aphasia Test* (BAT; Paradis & Libben, 1987). To pool effect sizes, we replaced the rows for these values with the average effect size and the average variance.

2.4.3. Trimming

As some scores in the pretest measures were perfect, we removed such rows as there was no possibility for improvement in the posttest or follow-up measures.

2.4.4. Multilevel model

As our data could include several observations from the same individual, or from several individuals from the same study (or both), effect sizes could not be considered independent. Dependent effect sizes are less informative than independent effect sizes. When effect sizes are correlated (that is, an improvement on one task in a particular individual is expected to correlate with an improvement on another task affected by the same language function), information obtained from one effect size overlaps with information obtained from the other effect size.

If this dependency is not taken into consideration, the amount of information is overestimated, and confidence intervals are underestimated, which results in an inflation of the risk of Type-I errors (e.g., Becker, 2000). To consider the dependency between effect sizes, we used a multilevel meta-analysis (e.g., Van den Noortgate et al., 2013). We considered the two following dependencies: 1) effect sizes drawn from one particular individual were assumed more similar than effect sizes drawn from another individual, and 2) effect sizes drawn from individuals within the same study (i.e., subject to similar study conditions) were assumed more similar than effect sizes drawn from individuals in different studies.

We compared the model fit indices of the plain (assuming all effect sizes could be treated as independent) and the two different multilevel models using a likelihood-ratio test. To do this, we used the *anova.rma* function in *metafor*(Viechtbauer, 2010). As our two-level model (modeling dependence within individuals) had significantly better model fit than our plain model, and our three-level model (modeling dependence within individuals and within studies) had better model fit than the two-level model, we continued our analyses with the three-level model (Table 2).

2.4.5. Assessment of bias

To assess and account for asymmetry in the distribution of outcomes, we conducted a close equivalent to a PET-PEESE analysis based on our three-level model. To do this, we added the standard error (*SE*) or variance (SE^2) for each outcome as a continuous predictor in the main analyses. In the precision-effect test (PET) and the precision-effect test with standard error (PEESE), outcomes are regressed on their SE/SE^2 in a weighted least-squares regression. If the association between *SE* and the size and direction of the outcomes is statistically significant in the PEESE test, this indicates a bias where studies of low precision systematically over- or underestimate effects. However, modeling studies suggest that the intercept (β_0) in the PEESE models tends to overcorrect the true effect, and that replacing the *SE* with the SE/SE^2 provides a better estimate of an effect for a hypothetical study with perfect statistical power.

We found a statistically significant association between the *SE* and the outcomes, $Q[1] = 143.60$, $p < 0.001$, such that outcomes with higher *SE* were associated with larger effects. Similarly, replacing the *SE* with the SE^2, the association was statistically significant, $Q[1] = 137.61$, $p <$ 0.001 (see Fig. 2). To improve the reliability of corrected estimates, PET-PEESE analyses were conducted only for analyses without moderators and with $k > 45$.

3. Results

3.1. Descriptive results

The final data file included 1,573 effect sizes from 85 individuals[2] and 40 studies. Of these effect sizes, 672 represented within-language effects, 701 cross-language effects, and 200 both, when there was only one treatment language. One hundred and sixteen effect sizes represented measurements with more than one treatment language. The number of effect sizes per participant ranged from 3 to 104. As the number of effect sizes varied considerably across participants, we report the number of participants (*n*) together with the number of effect sizes (*k*) in the subsequent analyses. The synthesized treatment outcomes for each study (i.e., the aggregated effect size across measures and participants) are presented in Fig. 3. For descriptive information about the participant- and task-related characteristics of the studies, as well as the results, see Tables S1 and S2.

[2] Note that the data file included 89 coded participants as four of them participated in two separate studies. In the reporting of *n* we, however, refer to the subset calculated from the 85 individuals.

3.2. Overall effects

We first estimated the overall effect size across all participants and all included measurements that had included one treatment language. We found a small to medium effect, $g = 0.38$ [0.29, 0.47], $p < 0.001$, $k = 1573$, $n = 85$, such that the participants performed better after the intervention than before. The test for heterogeneity was significant, $Q[1572] = 3069.93$, $p < 0.001$. After correcting for the asymmetry in the distribution of effect sizes using a PET-PEESE method, the estimated overall effect size was somewhat lower, $g = 0.23$ [0.14, 0.32], $p < 0.001$ (see Fig. 4).

3.2.1. Follow-up

To check maintenance of treatment effects, we also investigated the overall effect size for follow-up measures of the effect in the studies that had measured treatment outcomes also at a later stage. Because only a portion of the studies reported follow-up data, we examined those data irrespective of the time elapsed between the end of the intervention and the follow-up data collected. We found a small-to-medium positive outcome, $g = 0.42$ [0.26, 0.59], $p < 0.001$, $k = 131$, $n = 13$, $Q[130] = 220.48$, $p < 0.001$. The PET-PEESE correction led to a slightly lower estimate, $g = 0.33$ [0.15, 0.50], $p < 0.001$.

3.3. Within-language effects

After this, we limited the data to include pretests and posttests conducted in the same language as the treatment was performed in (see Fig. 4). We found a medium effect size, $g = 0.53$ [0.40, 0.66], $p < 0.001$, $k = 672$, $n = 82$, $Q[671] = 1540.17$, $p < 0.001$. Again, the PET-PEESE corrected effect was slightly lower, $g = 0.36$ [0.24, 0.48], $p < 0.001$. The control analysis produced similar estimates, $g = 0.59$ [0.44, 0.74], $p < 0.001$, $k = 467$, $n = 78$, $Q[466] = 1151.68$, $p < 0.001$, and the PET-PEESE corrected effect size, $g = 0.39$ [0.26, 0.53], $p < 0.001$.[3]

3.3.1. Follow-up

For follow-up measures, the synthesized within-language treatment outcome was large, $g = 0.70$ [0.42, 0.99], $p < 0.001$, $k = 58$, $n = 12$, $Q[57] = 97.97$, $p < 0.001$. The PEESE-corrected effect size was slightly smaller, $g = 0.63$ [0.26, 1.00], $p < 0.001$.

3.3.2. Domain

We then tested whether treatment outcomes were different based on language domain (production, comprehension, repetition) for pretest and posttest comparisons within the same language. We found a statistically significant difference, $Q[2] = 28.51$, $p < 0.001$, $k = 619$, $n = 79$. The treatment outcome was highest for production, $g = 0.62$ [0.48, 0.76], $p < 0.001$, $k = 412$, $n = 79$, somewhat lower for comprehension, $g = 0.41$ [0.26, 0.56], $p < 0.001$, $k = 187$, $n = 41$, and lowest for repetition (but note the lower *n* for repetition), $g = 0.35$ [0.10, 0.59], $p = 0.006$, $k = 20$, $n = 14$.

3.3.3. Treatment level

Following this, we conducted a moderation test to test whether the outcome differed depending on whether the linguistic level the treatment focused on (words, sentences, or discourse) was the same as (or different from) the level that the test measured. We found a statistically significant difference, $Q[1] = 21.32$, $p < 0.001$, $k = 670$, $n = 81$. The effects were larger within, $g = 0.61$ [0.47, 0.76], $p < 0.001$, $k = 477$, $n = 72$, than across levels, $g = 0.41$ [0.27, 0.56], $p < 0.001$, $k = 193$, $n = 46$.

[3] As some studies combined blocks of treatment in different languages or types, we checked whether the results are similar when only including the first treatment block. This control analysis produced similar estimates to those of the main analysis.

Table 2
Comparisons of model fit between the plain and the multilevel models.

		Model fit indices		Model comparison			Variance components	
Levels	Added levels	AIC	LogLIK	Models	LRT	p	σ_1^2	σ_2^2
1. One		3765.96	−1881.98					
2. Two	Individual	3240.42	−1618.21	1 vs. 2	527.54	<0.001	0.07	
3. Three	Study	3235.49	−1614.74	2 vs. 3	6.95	0.008	0.05	0.03

Note. AIC = Akaike Information Criterion; LogLik = Log-Likelihood; LRT = Likelihood-Ratio Test. The Likelihood-Ratio test statistic is tested against a chi-square distribution with 1 degree of freedom.

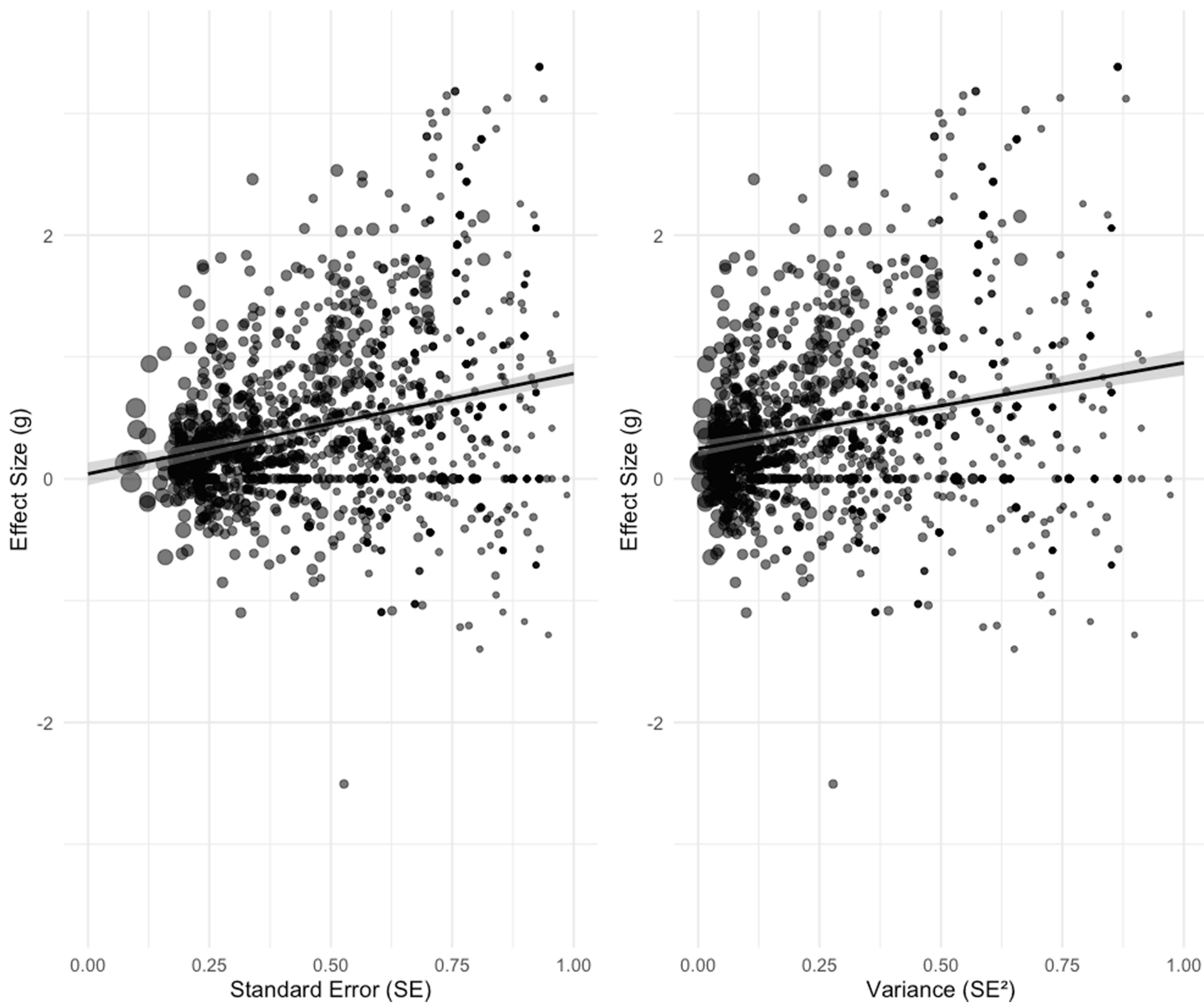

Fig. 2. Outcomes regressed on their standard error (left panel) and on their variance (right panel).

3.3.3.1. Within specific level. We continued with a more specific analysis of differences in outcomes when the linguistic level of treatment was the same as the level the test measured (word or sentence; there were no observations with discourse). We found no statistically significant differences between these levels, $Q[1] = 0.26$, $p = 0.609$, $k = 429$, $n = 71$. There was a medium effect at the level of word, $g = 0.52$ [0.35, 0.69], $p < 0.001$, $k = 344$, $n = 58$, and at the level of sentence, $g = 0.60$ [0.36, 0.83], $p < 0.001$, $k = 85$, $n = 15$ (but note the lower n for sentence level).

3.3.3.2. Between specific levels. After this, we limited analyses to situations in which the linguistic level the treatment focused on (words, sentences, or discourse) was different from the level that the test measured to look at the difference between higher and lower level. We found no statistically significant difference between the levels, $Q[3] = 7.56$, $p = 0.056$, $k = 378$, $n = 51$. There was a medium effect for "two steps higher", $g = 0.62$ [0.37, 0.87], $p < 0.001$, $k = 27$, $n = 8$, somewhat smaller effects for "one step higher", $g = 0.33$ [0.23, 0.42], $p < 0.001$, $k = 180$, $n = 40$, and for "one step lower", $g = 0.31$ [0.20, 0.42], $p < 0.001$, $k = 161$, $n = 17$, and a non-significant effect for "two steps lower" (but with a very small n), $g = 0.18$ [−0.04, 0.39], $p = 0.104$, $k = 10$, $n = 2$.

3.3.4. Trained items

We then investigated whether treatment outcomes differed between trained and untrained items within the same tasks. We found a statistically significant difference, $Q[1] = 109.66$, $p < 0.001$, $k = 157$, $n = 50$, such that effects were large for trained, $g = 1.14$ [0.91, 1.36], $p < 0.001$, $k = 72$, $n = 50$, and small-to-medium for untrained, $g = 0.36$ [0.14, 0.59], $p = 0.002$, $k = 85$, $n = 50$, items.

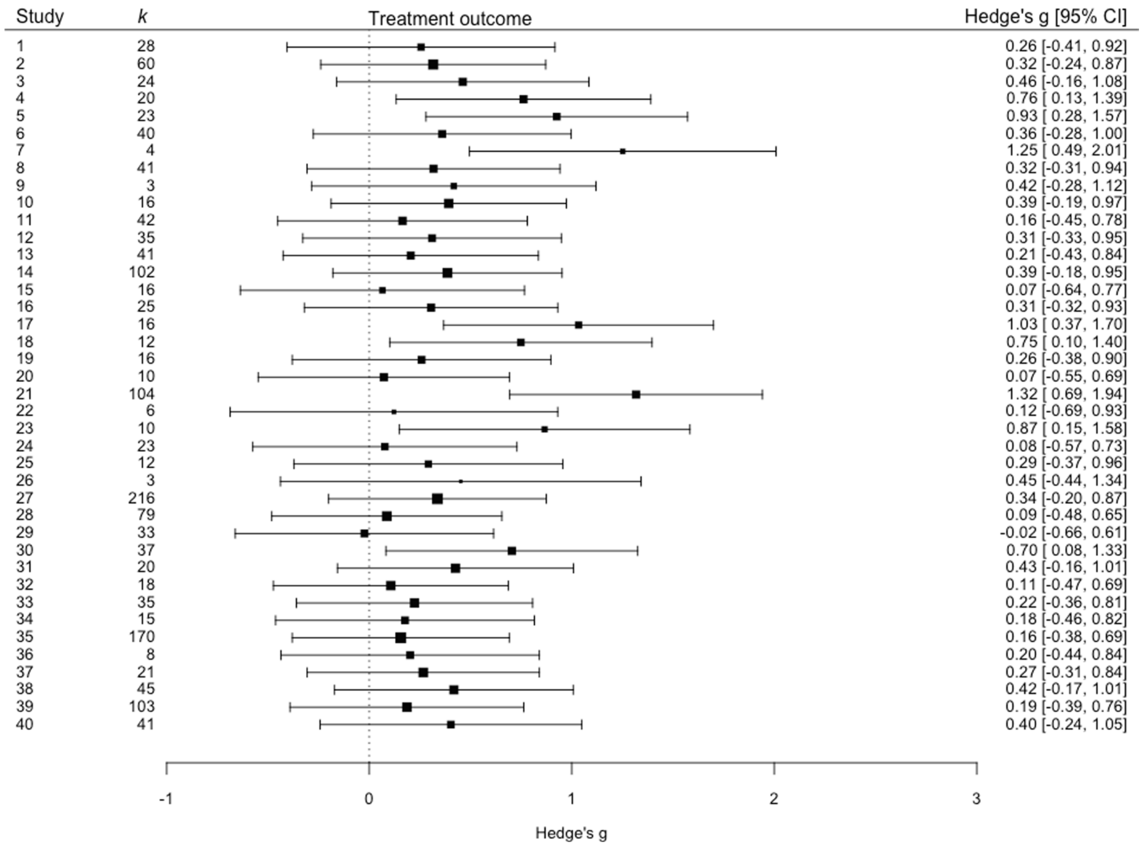

Fig. 3. Forest plot over effect size estimates per study and synthesized effect.

3.3.5. Treatment language

To test whether the treatment language affected the outcome, we analyzed the difference between treatment provided in L1 and Ln. We did not find a statistically significant difference, $Q[1] = 0.27$, $p = 0.605$, $k = 672$, $n = 82$, with effect sizes being relatively similar for L1, $g = 0.55$ [0.40, 0.69], $p < 0.001$, $k = 260$, $n = 41$, to those for Ln, $g = 0.52$ [0.38, 0.65], $p < 0.001$, $k = 412$, $n = 51$. For within-level effects, we did not find a statistically significant difference $Q[1] = 0.21$, $p = 0.648$, $k = 477$, $n = 75$, between L1, $g = 0.58$ [0.42, 0.73], $p < 0.001$, $k = 182$, $n = 38$, and Ln, $g = 0.55$ [0.41, 0.69], $p < 0.001$, $k = 295$, $n = 46$. For across-levels, we did not find a statistically significant difference, either, $Q[1] = 0.13$, $p = 0.719$, $k = 193$, $n = 49$. The effects for L1, $g = 0.41$ [0.23, 0.59], $p < 0.001$, $k = 78$, $n = 22$, were relatively similar to those in Ln, $g = 0.44$ [0.29, 0.60], $p < 0.001$, $k = 115$, $n = 29$.

3.3.6. Age of acquisition

We then tested whether within-language treatment effects were dependent on the age of acquisition of the treated language (see Fig. 5). We found a statistically significant difference, $Q[3] = 11.05$, $p = 0.011$, $k = 649$, $n = 85$. For the first language, the treatment effect was $g = 0.56$ [0.41, 0.72], $p < 0.001$, $k = 300$, $n = 45$. For early childhood, the treatment effect was $g = 0.63$ [0.42, 0.84], $p < 0.001$, $k = 50$, $n = 11$ (but note the lower n for early childhood). For later childhood, the treatment effect was $g = 0.35$ [0.18, 0.52], $p < 0.001$, $k = 237$, $n = 22$. For adulthood, the treatment effect was $g = 0.70$ [0.45, 0.95], $p < 0.001$, $k = 62$, $n = 11$.

3.3.7. Language proficiency

After this, we investigated if language proficiency in the treated language was associated with treatment outcomes in a within-language context. We found no significant difference between the proficiency groups, $Q[2] = 1.06$, $p = 0.588$, $k = 659$, n = 79. Treatment effects were of medium size in the high proficiency group, $g = 0.53$ [0.40, 0.66], $p < 0.001$, $k = 570$, $n = 67$, and only somewhat lower in the medium proficiency group, $g = 0.48$ [0.28, 0.67], $p < 0.001$, $k = 59$, $n = 11$, and in the low proficiency group, $g = 0.39$ [0.05, 0.74], $p = 0.025$, $k = 30$, $n = 4$, which included very few participants.

3.3.8. Aphasia severity

After this, we investigated whether treatment outcomes were associated with aphasia severity. Aphasia severity was coded from 1 = mild to 5 = severe and treated as a linear predictor. We found no statistically significant association between severity and the treatment outcome, $g = 0.04$ [−0.02, 0.10], $p = 0.169$, $k = 488$, $n = 54$, $Q[1] = 1.89$, $p = 0.169$.

3.4. Cross-language effects

We then investigated the treatment effect for all studies across languages (see Fig. 4). We found a small but statistically significant effect, $g = 0.29$ [0.18, 0.41], $p < 0.001$, $k = 701$, $n = 77$, $Q[700] = 1265.99$, $p < 0.001$. The PET-PEESE corrected effect was weaker but still statistically significant, $g = 0.14$ [0.03, 0.26], $p = 0.015$. In the control analyses, the effect size estimate was $g = 0.35$ [0.22, 0.48], $p < 0.001$, $k = 539$, $n = 74$, $Q[538] = 1008.81$, $p < 0.001$, and the PET-PEESE corrected effect size, $g = 0.19$ [0.06, 0.31], $p = 0.003$.[4]

3.4.1. Follow-up

For follow-up measures, the synthesized cross-language treatment outcome did not differ statistically from null, $g = 0.13$ [−0.13, 0.39], p

[4] As some studies combined blocks of treatment in different languages or types, we checked whether the results are similar when only including the first treatment block. The results of this control analysis showed only a minimal difference to the main results.

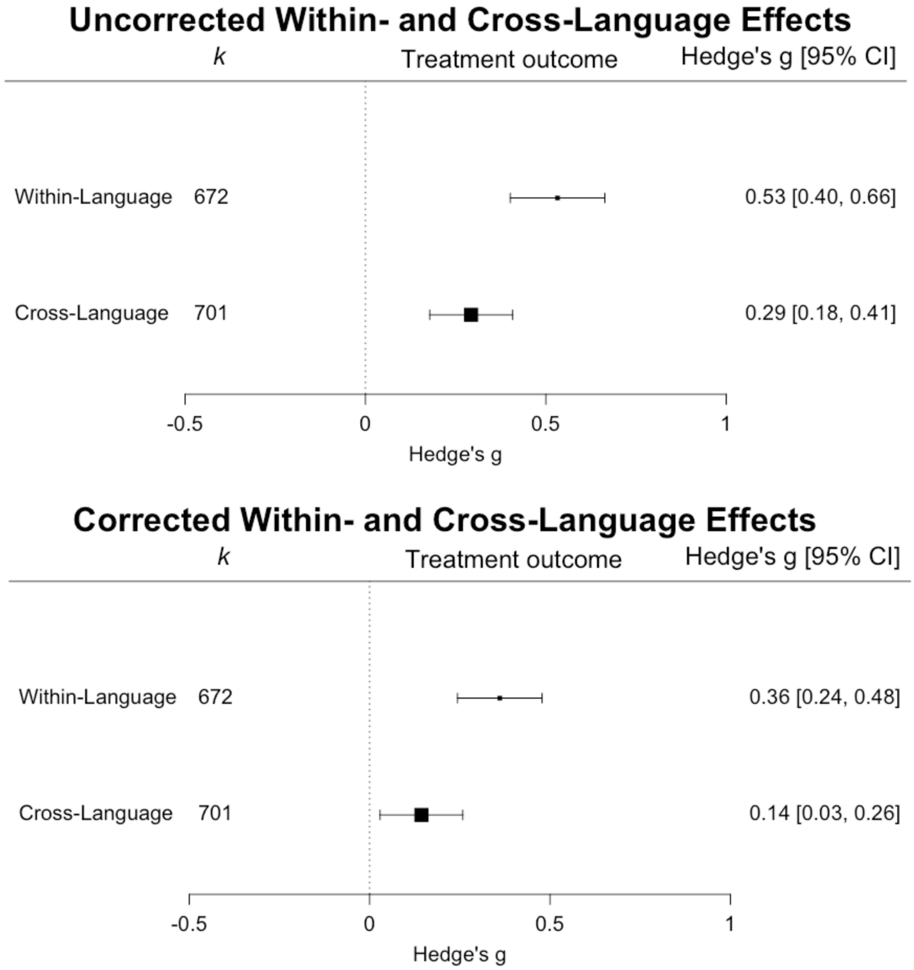

Fig. 4. Uncorrected (upper panel) and corrected (lower panel) estimates for treatment effects within and across languages.

$= 0.318$, $k = 61$, $n = 11$, $Q[60] = 55.14$, $p = 0.654$. The PEESE-corrected effect size was even smaller, $g = -0.01$ [-0.27, 0.24], $p = 0.917$.

3.4.2. Domain

We again tested whether cross-language treatment outcomes were different based on language domain (production, comprehension, repetition), after removing 12 effect sizes representing both comprehension and production (e.g., translation). We did not find a statistically significant difference, $Q[2] = 0.61$, $p = 0.737$, $k = 627$, $n = 77$. The treatment outcomes were relatively similar for production, $g = 0.29$ [0.17, 0.41], $p < 0.001$, $k = 384$, $n = 77$, comprehension, $g = 0.32$ [0.19, 0.45], $p < 0.001$, $k = 227$, $n = 39$, and repetition, $g = 0.35$ [0.09, 0.61], $p = 0.008$, $k = 16$, $n = 11$ (but note the lower n for repetition).

3.4.3. Treatment level

We then conducted a moderation test to study whether the outcome differed depending on whether the linguistic level the treatment focused on was the same as (or different from) the level that the test measured. We found no statistically significant difference, $Q[1] = 0.00$, $p = 0.945$, $k = 698$, $n = 76$. The effects were almost identical for within, $g = 0.29$ [0.17, 0.41], $p < 0.001$, $k = 532$, $n = 70$, and across, $g = 0.29$ [0.16, 0.42], $p < 0.001$, $k = 166$, $n = 42$, levels.

3.4.3.1. Within specific level. We then analyzed the possible differences in outcomes when the linguistic level of treatment was the same as the level the test measured. Due to limited data, we included only the levels of word and sentence (there were no treatment outcomes available for discourse). We found no statistically significant differences between these levels, $Q[1] = 0.36$, $p = 0.547$, $k = 506$, $n = 69$. There were similar effects for word, $g = 0.26$ [0.10, 0.43], $p = 0.002$, $k = 347$, $n = 57$, and sentence, $g = 0.34$ [0.13, 0.55], $p = 0.002$, $k = 159$, $n = 14$ (but note the lower n for sentence level).

3.4.3.2. Between specific levels. After this, we limited analyses to situations in which the linguistic level the treatment focused on (e.g., words or sentences) was different from the level that the test measured to look at the difference between higher and lower levels. We found no statistically significant difference between the levels, $Q[3] = 4.51$, $p = 0.211$, $k = 165$, $n = 42$. The effects were relatively similar for "one step lower", $g = 0.41$[0.22, 0.61], $p < 0.001$, $k = 68$, $n = 16$, "one step higher", $g =$ 0.26 [0.09, 0.42], $p = 0.003$, $k = 78$, $n = 35$, "two steps lower", $g = 0.28$ [-0.08, 0.64], $p = 0.131$, $k = 4$, $n = 2$, and for "two steps higher", $g =$ 0.46 [0.13, 0.80], $p = 0.007$, $k = 15$, $n = 7$, the latter two categories with very few observations.

3.4.4. Translation equivalents of trained items

We then investigated whether treatment outcomes differed between translation equivalents of trained and untrained items within the same tasks. We found a statistically significant difference, $Q[1] = 7.67$, $p =$ 0.006, $k = 147$, $n = 45$. Effects were larger for trained, $g = 0.57$ [0.33, 0.80], $p < 0.001$, $k = 67$, $n = 45$, than for untrained, $g = 0.35$ [0.12, 0.59], $p = 0.003$, $k = 80$, $n = 45$, items.

3.4.5. Treatment language

To test whether the treatment language affected the outcome, we analyzed the difference between L1 and Ln. We found no statistically

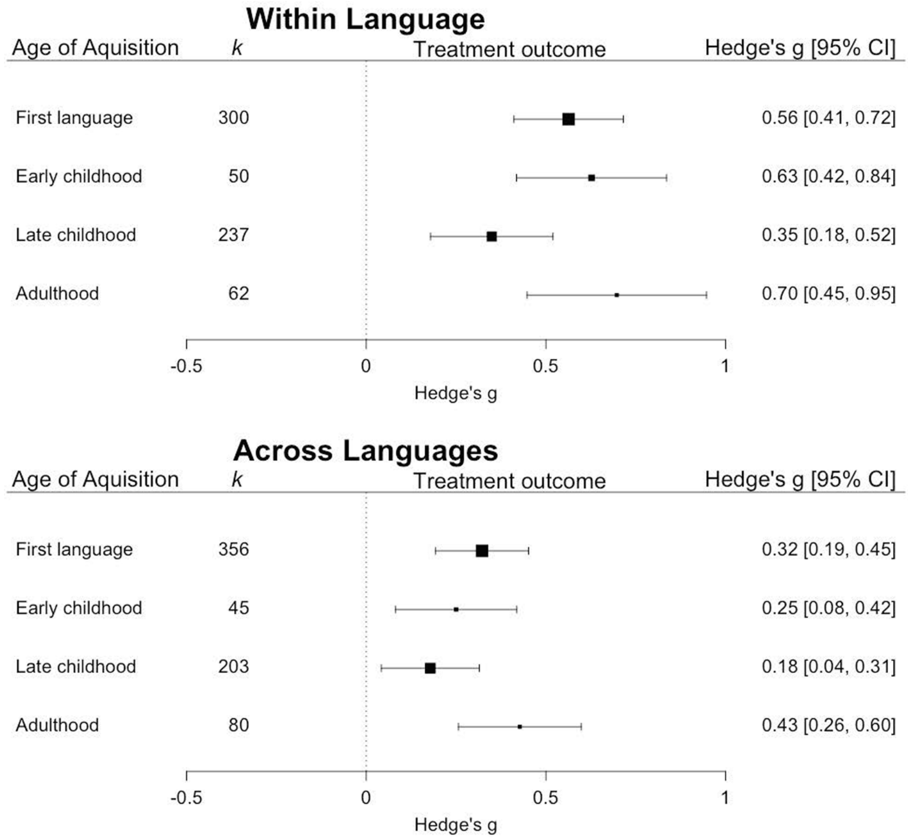

Fig. 5. Estimates for treatment effects by age of acquisition of the treated language for within-language effects (upper panel) and across-language effects (lower panel).

significant difference, $Q[1] = 0.38$, $p = 0.540$, $k = 701$, $n = 77$. The effect sizes were similar with $g = 0.31$ [0.18, 0.44], $p < 0.001$, $k = 261$, $n = 38$, for L1 and $g = 0.28$ [0.16, 0.40], $p < 0.001$, $k = 440$, $n = 50$. for Ln. Similarly, we found no statistically significant difference between L1 and Ln for within-level, $Q[1] = 0.40$, $p = 0.528$, $k = 532$, $n = 70$, nor for across-levels, $Q[1] = 0.04$, $p = 0.851$, $k = 166$, $n = 42$.

3.4.6. *Age of acquisition*

We studied whether age of acquisition of the treatment language or that of the test language was associated with cross-language effects. We found a statistically significant association between age of acquisition of the treatment language and cross-language treatment outcomes $Q[3] = 9.90$, $p = 0.020$, $k = 684$, $n = 75$, where the largest outcomes were observed when the treated language had been acquired in adulthood, $g = 0.43$ [0.26, 0.60], $p < 0.001$, $k = 80$, $n = 11$, followed by effects for first-language acquisition $g = 0.32$ [0.19, 0.45], $p < 0.001$, $k = 356$, $n = 42$, early childhood $g = 0.25$ [0.08, 0.42], $p = 0.004$, $k = 45$, $n = 11$, and finally for later in childhood, $g = 0.18$ [0.04, 0.31], $p = 0.010$, $k = 203$, $n = 22$. There was, however, no effect of age of acquisition of the test language in cross-language contexts, and $Q[3] = 1.20$, $p = 0.752$, $k = 701$, $n = 77$.

3.4.7. *Language proficiency*

We then tested if cross-language generalization effects depend on the proficiency relationship of the languages. We found no statistically significant difference, $Q[2] = 0.44$, $p = 0.803$, $k = 689$, $n = 74$. Effects were similar for "same", $g = 0.28$ [0.17, 0.40], $p < 0.001$, $k = 531$, $n = 54$, and for "tr_higher", $g = 0.31$ [0.14, 0.47], $p < 0.001$, $k = 59$, $n = 12$, and for "tr_lower", $g = 0.26$ [0.11, 0.40], $p < 0.001$, $k = 99$, $n = 15$.

3.4.8. *Language distance between treatment vs. test language*

To test whether cross-language generalization effects depend on the similarity of the languages, we coded language distance on a 10-step scale (0–9) with higher values representing higher similarity. We used this variable as a linear predictor for the pre-post effect of treatment. We found no significant association, $g = 0.01$ [−0.04, 0.05], $p = 0.771$, $k = 719$, $n = 77$, $Q[1] = 0.08$, $p = 0.771$.

3.4.9. *Aphasia severity*

We then investigated whether treatment outcomes were associated with aphasia severity. Again, we found no statistically significant association between severity and the treatment outcome, $g = -0.00$ [−0.05, 0.05], $p = 0.929$, $k = 459$, $n = 49$, $Q[1] = 0.01$, $p = 0.929$.

3.5. *Treatment in both languages*

Finally, we investigated the treatment outcome in 116 instances and seven participants where the treatment had been given in both languages within the same treatment block. We found no effect for this small number of cases, $g = 0.12$ [−0.01, 0.25], $p = 0.079$, $k = 116$, $n = 7$, $Q[115] = 108.74$, $p = 0.647$. The PEESE correction did not change this finding, $g = -0.02$ [−0.21, 0.17], $p = 0.827$.

4. Discussion

Our systematic review yielded 40 studies reporting on 85 multilingual people with aphasia who received language intervention in one or more of their languages and were assessed in the language used for treatment as well as in their other language(s). We conducted a meta-analysis of the 1573 effect sizes extracted from these studies. We address our results with respect to our research questions.

4.1. Is language intervention with multilingual people with aphasia beneficial?

Our findings indicate that the answer to our first question is yes: Overall, participants' performance was higher in the post-treatment measurements as compared to the pre-treatment ones. Also as predicted, treatment effects were stronger within languages than across languages. That is, larger effect sizes emerged for treatment effects in the treated language than in the untreated language(s), for which we found very small overall effects when correction for bias was applied. In addition, whereas a relatively small number of follow-up measurements were reported ($n = 12$ for within-language and 11 for cross-language), those revealed maintenance of treatment effects, but only in the treated languages; cross-language follow-up effects did not show a significant effect.

These results are in line with previous studies reporting positive treatment outcomes for multilingual people with aphasia, with stronger within- than between-language effects and no negative effects for the untreated language (e.g., Faroqi-Shah et al., 2010). Finding positive treatment outcomes for multilingual people with aphasia is significant because many of the people with aphasia seen for intervention worldwide are multilingual, whereas the majority of the intervention approaches used by speech-language therapists have been developed and tested on monolingual people with aphasia (e.g., Ansaldo et al., 2010; Grasso et al., 2019). Moreover, finding cross-language generalization would be important because many multilingual people with aphasia have access to treatment in only one of their languages, often the language of the majority in the country where they reside at the time of the stroke (Centeno, 2009; Norvik et al., 2022; Sandberg et al., 2019). Yet, findings for cross-language generalization have been elusive, with some studies reporting effects and others not, which can explain the smaller effects we found in our meta-analysis.

We note that treatment outcomes were higher for measures of language production than of language comprehension for the treated languages. This is consistent with the fact that most language treatment approaches for aphasia typically target production skills, at least in the sample included in this systematic review. Intervention approaches included in studies with multilingual people with aphasia, such as Semantic Features Analysis (SFA, e.g., Bihovsky et al., 2023), Verb Network Strengthening Therapy (VNeST, e.g., Lerman et al, 2019), Script Training (e.g., Grasso et al., 2019), target lexical-retrieval production, as anomia (word-finding difficulties) is a major deficit in all types of aphasia. Although these treatments (e.g., SFA, VNeST) target also semantic processing, and thus could be expected to affect comprehension processing as well as production, we still observed stronger effects for production tasks. This may be related to the observation that relatively few individual studies measured language comprehension as a direct treatment outcome, though many studies used the *Western Aphasia Battery* (WAB; Kertesz, 1982) or the *Bilingual Aphasia Test* (BAT; Paradis & Libben, 1987) as outcome measures – aphasia batteries that include comprehension subtests. A dissociation between production and comprehension outcomes in the generalization of treatment is informative for clinicians who aim to maximize treatment effects not only in the treated language but also in the untreated languages. It can also explain some of the mixed results for cross-language generalization reported in the literature.

4.2. What variables affect treatment efficacy?

With respect to our second research question, several variables emerged as contributing to the efficacy of the treatment. We address the treatment-related variables, followed by the multilingualism-related variables, below.

The treatment-related variables we examined included the linguistic levels targeted and tested, and whether the outcome measures consisted of items that were trained during the treatment or not. One, we found stronger treatment effects when the linguistic level of the intervention (e.g., words vs. sentences) was the same as the level measured. This was true for the within-language analyses but not for the cross-language analyses. However, we note that in many interventions, more than one linguistic level is targeted so it is possible that even though the main goal of an intervention is to improve sentence construction, for example, individual words are targeted as well, and, vice versa, in treatment approaches that targeted single words, some activities may elicit language beyond the word level. Consistently, we found significant effects for both within- and between-levels effects, as well as no significant difference between levels in the cross-language effects. This is not entirely surprising, and perhaps reassuring, because the aim of aphasia treatment is to generalize beyond the items and structures targeted. Thus, improved performance at post-treatment testing compared to pre-treatment baselines, regardless of the linguistic levels targeted during the treatment, is an encouraging finding.

Two, when we examined outcomes for trained vs. untrained items, trained items – as predicted – showed greater benefit from treatment than untrained items. The comparison of treated vs. untreated material has been at the focus of many intervention studies. In treatment studies that target specific items, such as those using SFA or VNeST and those targeting naming abilities, it is typical to find greater improvements in the trained items than in items that were not directly practiced during therapy (e.g., Boyle, 2010). Untrained items within the treated language have consistently been reported to benefit less from treatments that target specific items (e.g., Edmonds & Kiran, 2006; Kiran & Roberts, 2010). Indeed, generalization of treatment-related improvements to untrained items has been difficult to achieve in aphasia treatment more generally (e.g., Best et al., 2013; Fillingham et al., 2006). Thus, our findings of stronger effects for directly trained items are commensurate with what may be expected from previous treatment studies with monolingual and multilingual participants. Also consistent with previous reports (e.g., Kohnert, 2004; Marte et al., 2023), we found that translation equivalents of trained items benefited more from the treatment than other words in the untreated language.

The three multilingualism-related variables that we were able to examine in the current dataset were age of language acquisition, language proficiency of the treated and untreated languages, and language distance. The only variable that yielded a significant effect was age of language acquisition of the treated language. That is, treatment in languages that were acquired from birth showed better efficacy than treatment in languages that were learned in early childhood, which in turn showed better response to treatment than languages that were learned in later childhood. This was true for both within-language and cross-language results. The finding of better outcomes for treatment in a language acquired from birth is consistent with previous reports of better post-stroke recovery of the first-acquired language (e.g., Faroqi-Shah et al., 2010; Kuzmina et al., 2019; Goral, 2022). Although the better recovery of the first-acquired language has not been found consistently in the literature on multilingual people with aphasia, it has been reported in many instances, and as early as in Ribot (1881) who proposed that the first-acquired language would be more resilient to the effect of aphasia. Yet, quite a few cases have been reported to the contrary of this prediction, starting with Pitres (1895) who countered Ribot's law with the prediction that the language most used at the time of the stroke, not the one acquired first, would be more likely to recover in aphasia. In fact, our findings did not extend to the age of acquisition of the tested language, qualifying the early acquisition effect.

In case studies or case series of treatment studies, isolating the effects of individual variables such as age of language acquisition can be challenging given the heterogeneity of the multilingual population enrolled in such studies and the likely interaction of a number of variables with age of language acquisition, such as language use and language proficiency. The meta-analysis employed here allowed us to demonstrate that the effect of age of language acquisition was significant, independent of language proficiency. Yet, language proficiency and language use no doubt play a role in patients' response to treatment.

For example, language exposure and use can explain the somewhat unexpected finding obtained in our analyses: whereas treatment that was administered in an early-acquired language was more efficacious than treatment that was administered in a language that was acquired later in childhood, when the treatment was administered in a non-L1 that was acquired in adulthood, treatment effects were strong. In those 11 cases in our sample who were treated in a language they learned in adulthood, the late-learned language had become the main language of the participants as they were immersed in that language and used it consistently prior to the stroke. They all used their late-learned language extensively (although not all had comparable proficiency in that language and in their earlier-acquired languages). It is therefore possible that language use modulates the effect of age of acquisition. The role of language use has emerged as an important variable in bilingualism (e.g., Peñaloza et al., 2020). We were unable to separately examine the effect of language use on treatment outcome in our current analyses due to inconsistent reports about current and past use of each of the languages of the participants included in this review, but it is possible that language immersion and use were the drive behind the effect of the later age.

In contrast to the effect of age of acquisition, language proficiency did not appear to affect within-language or cross-language effects. The finding of no language proficiency effect for cross-language treatment outcome is of interest because the role of language proficiency has been debated in the literature on cross-language generalization (e.g., Ansaldo & Ghazi Saidi, 2014; Conner et al., 2018; Goral & Lerman, 2020; Knoph et al., 2017). In a recent study, Peñaloza et al. (2020) demonstrated that pre-stroke proficiency predicted post-stroke lexical-semantic abilities in bilingual people with aphasia. Pre-stroke language proficiency and residual abilities post-stroke have been hypothesized to affect the processes assumed to underlie cross-language treatment generalization. For example, processes of spreading activation among lexical items should lead to generalization from a less-proficient language to a more proficient one, in accordance with models of the bilingual lexicon (e.g., Kiran et al., 2013; Kroll et al., 2010). This is because of the assumed stronger lexical connections from L2 words to L1 words than in the opposite direction (Kiran et al., 2013). At the same time, the need to inhibit a more-proficient language while activating a less-proficient one can lead to a lingering suppression of a more-proficient language following treatment in a less-proficient language (Goral et al., 2013). In our systematic review, the vast majority of the study participants had medium or high proficiency in all their languages so the finding of no significant language proficiency effect may be expected. It is possible that the mixed results among the individual studies, with some reporting cross-language effects from a weaker to a stronger language and some in the other direction, and with both processes – spreading activation and inhibition – at play, resulted in a null effect of proficiency in this meta-analysis.

Finally, and consistent with previous reviews that found no effect of language distance on whether cross-language generalization was observed (Ansaldo & Ghazi Saidi, 2014; Faroqi-Shah et al., 2010), we found no effect of language distance in our analyses. We note that even those reviews that did not find a significant effect of language distance do point to the role of language similarities, especially that of cognates (Ansaldo & Ghazi Saidi, 2014; Kohnert, 2004). It is possible that for treatments that focus on specific lexical items, lexical similarity is a more relevant variable than overall language distance. For treatment approaches that target other linguistic processes, language distance may be a variable that contributes to cross-language effects.

Additional variables that have been implicated in the treatment efficacy literature include aphasia severity and aphasia type. Aphasia severity did not emerge as a predictor in within- or cross-language effects in our analyses. We did not include aphasia type as a variable in our analyses. This is because the classification of people with aphasia into aphasia types has received much criticism and discussion in recent years (e.g., Ardila, 2010) and many studies included in this review did not choose to classify their participants according to classification methods, such as the syndrome approach.

4.3. Limitations and future directions

Our search of four major databases using all relevant key terms yielded 40 studies, many of single participants or small groups, for inclusion in this systematic review and meta-analysis. It is possible that the relatively small sample size obtained contributed to some of the null effects observed. Nevertheless, the sample yielded 1573 observations, which offered considerable statistical power. Moreover, in some comparisons (e.g., proficiency), the number of cases compared were uneven, so the interpretation of the results should be done with caution. For example, smaller sample sizes in some of the conditions may lead to less precise estimates and increase the risk of false positive or negative findings as well as limit the generalizability of the findings. It is possible that the difference in the treatment effect in these instances is at least in part due to the size of the groups.

We observed asymmetry in the distribution of outcomes, indicating that studies with lower precision tended to report larger positive effects than those with higher precision. Such asymmetries may be observed due to reporting or publication biases. To account for this asymmetry, we used a correction close to the PET-PEESE method that estimates what the effect size would be in the absence of bias. It should be noted that many of the present moderator analyses were not corrected for bias. Furthermore, as none of the methods to correct for publication bias are perfect, our adjusted effect size values should be treated with some caution.

As well, the heterogeneity inherent to the target population likely contributed to the small effects of cross-language treatment generalization. Furthermore, even though we reviewed the studies carefully in an attempt to extract as much information about the participants, variation in how information pertaining to language proficiency, age of acquisition, and especially language use is report, limited our ability to analyze the contribution of some of these variables to treatment efficacy in multilingual people with aphasia. It may be beneficial for the field if researchers increased the consistency with which they report critical information about language exposure and use across the lifespan of their participants. Several questionnaires have been reported across studies but no gold standard for collecting and reporting this information has emerged. Determining pre-stroke proficiency has been acknowledged as a challenge, given that only subjective information is available (Lerman et al., 2020). Peñaloza et al. (2020), who demonstrated the importance of pre-stroke proficiency, calculated pre-stroke proficiency taking into account language use history, education history, and lifelong exposure. In many previous studies, only part of this information was utilized to determine proficiency. Understanding the role of relative proficiency is especially critical given recent discussions in the literature and theoretical and practical implications. If treating one language, for example a more proficient or a more used one, could lead to better outcomes, such findings can guide clinical decisions.

As in any treatment study in aphasia, improvement following treatment can be the result of the treatment but other factors may contribute to change, including spontaneous recovery in early stages post stroke and general response to the engagement during treatment. In the cases included here, the majority of the individuals were in the chronic phase post stroke and little spontaneous recovery could have been expected. As well, several of the studies included here applied measures of control, including tasks that were unrelated to the treatment and were not expected to change following the intervention (and were thus not part of the meta-analysis).

Finally, great variation in both the treatment approaches used and in the outcome measures reported in the various studies also contributed to variability in the sample. Indeed, because of the relatively small number of studies included here, we were unable to examine treatment effects separately for specific intervention approaches or approaches that employed specific activities. This heterogeneity of treatment approaches is typical of treatment studies in aphasia so in this respect the present sample does represent the literature and our results should be generalizable. With the growing interest in multilingualism seen in the literature, we predict that additional aphasia intervention studies will continue to inform us about the variables that can help maximize within- and cross-language treatment generalization.

4.4. Conclusion

Our systematic review and a meta-analysis of studies that report treatment outcomes in multilingual people with aphasia suggests that treatment is beneficial for the language in which the treatment is conducted and also, though to a lesser extent, to the untreated language(s). On the basis of the results of this meta-analysis we can conclude that treating multilingual people with aphasia in one language may to a small extent generalize to their other languages, especially following treatment in an early-acquired language and in a later-learned language that became the language of immersion.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was partly supported by the Research Council of Norway through its Centres of Excellence funding scheme [223265]. We thank Ekaterina Kuzmina for her work on the earlier stage of this project and are grateful for the help of Sarah Cameron and Priscilla Saldivia with the initial data extraction.

Author contribution: MG contributed to the design of the study, search process, data extraction and coding, and the writing and editing of the manuscript. MN contributed to the design of the study, search process, data coding, and editing of the manuscript. JA conducted the statistical analyses and contributed to the writing and editing of the manuscript. IA contributed to data coding and the editing of the manuscript. ML contributed to the design of the study, search process, data extraction and coding, and the editing of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2023.105326.

References

* Provides the key that these papers were included in the systematic review/meta-analysis.

Albert, M. L., & Obler, L. K. (1978). *The bilingual brain: neuropsychological and neurolinguistic aspects of bilingualism*. Academic Press.

Ansaldo, A. I., & Ghazi Saidi, L. (2014). Aphasia therapy in the age of globalization: cross-linguistic therapy effects in bilingual aphasia. *Behavioural Neurology, 2014*, 1–10. https://doi.org/10.1155/2014/603085

*Ansaldo, A. I., Ghazi Saidi, L., & 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

Ardila, A. A. (2010). A proposed reinterpretation and reclassification of aphasic syndromes. *Aphasiology, 24*(3), 363–394. https://doi.org/10.1080/02687030802553704

Becker, B.J. (2000). Multivariate meta-analysis. In H.E.A. Tinsley & E.D. Brown (eds.), Handbook of applied statistics and mathematical modeling (pp. 499−525). Academic Press.

Best, W., Greenwood, A., Grassly, J., Herbert, R., Hickin, J., & Howard, D. (2013). Aphasia rehabilitation: Does generalisation from anomia therapy occur and is it predictable? A case series study. *Cortex, 49*(9), 2345–2357. https://doi.org/10.1016/j.cortex.2013.01.005

Bihovsky, A., Ben-Shachar, M., & Meir, N. (2023). Semantic Feature Analysis (SFA) treatment in L1 and L2 in bilingual aphasia: Effects of cognitive and language factors. *Aphasiology*. https://doi.org/10.1080/02687038.2023.2226359

Borenstein, M., Hedges, L.V., Higgins, J.P.T, & Rothstein, H.R. (2011). *Introduction to Meta-Analysis*. John Wiley & Sons Ltd.

Boyle, M. (2010). Semantic feature analysis treatment for aphasic word retrieval impairments: What's in a name? *Topics in Stroke Rehabilitation, 17*(6), 411–422. https://doi.org/10.1310/tsr1706−411

Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *The Cochrane Database of Systematic Reviews, 2016*(6), CD000425. https://doi.org/10.1002/14651858.CD000425.pub4

Centeno, J. G. (2009). Issues and principles in service delivery to communicatively impaired minority bilingual adults in neurorehabilitation. *Seminars in Speech and Language, 30*(3), 139–152. https://doi.org/10.1055/s−0029−1225951

Cherney, L. R., & Robey, R. R. (2008). Aphasia treatment: Recovery, prognosis, and clinical effectiveness. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related neurogenic communication disorders* (5th ed., pp. 186–202). Lippincott Williams & Wilkins.

Conner, P. S., Goral, M., Anema, I., Borodkin, K., Haendler, Y., Knoph, M., Mustelier, C., Paluska, E., Melnikova, Y., & Moeyaert, M. (2018). The role of language proficiency and linguistic distance in cross-linguistic treatment effects in aphasia. *Clinical Linguistics & Phonetics, 32*(8), 739–757. https://doi.org/10.1080/02699206.2018.1435723

*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)

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

Fillingham, J. K., Sage, K., & Lambon Ralph, M. A. (2006). The treatment of anomia using errorless learning. *Neuropsychological Rehabilitation, 16*(2), 129–154. https://doi.org/10.1080/09602010443000254

Goral, M. (2022). What can aphasia tell us about how the first-acquired language is instantiated in the brain? *Languages, 7*(4), 283. https://doi.org/10.3390/languages7040283

Goral, M., & Lerman, A. (2020). Variables and mechanisms affecting response to language treatment in multilingual people with aphasia. *Behavioral Sciences (Basel, Switzerland), 10*(9), 144. https://doi.org/10.3390/bs10090144

*Goral, M., Naghibolhosseini, M., & Conner, P. S. (2013). Asymmetric inhibitory treatment effects in multilingual aphasia. *Cognitive Neuropsychology, 30*(7–8), 564–577. https://doi.org/10.1080/02643294.2013.878692

*Grasso, S. M., Cruz, D. F., Benavidez, R., Peña, E. D., & Henry, M. L. (2019). Video-implemented script training in a bilingual Spanish-English speaker with aphasia. *Journal of Speech, Language, and Hearing Research: JSLHR, 62*(7), 2295–2316. https://doi.org/10.1044/2018_JSLHR-L−18−0048

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

Hallowell, B. & Chapey, R. (2008). Introduction to Language Intervention Strategies in Adult Aphasia. In R. Chapey (ed.), *Language Intervention Strategies in Aphasia and Related Neurogenic Communication Disorders* (5th ed., pp. 3−19). Lippincott Williams & Wilkins.

Hammarström, H., Forkel, R., Haspelmath, M., & Bank, S. (2022). *Glottolog 4.7*. Max planck institute for evolutionary anthropology. https://doi.org/10.5281/zenodo.7398962 (Available online at http://glottolog.org,).

Kertesz, A. (1982). *The Western aphasia battery*. Grune & Stratton.

Kiran, S. & Gray, T. (2018). Understanding the nature of bilingual aphasia: Diagnosis, assessment and rehabilitation. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (eds.), *Bilingual Cognition and Language: The state of the science across its subfields* (pp. 371−400). John Benjamins.

*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), 298–309. https://doi.org/10.1044/1058−0360(2013/12−0085)

*Knoph, M. I. N., Lind, M., & Simonsen, H. G. (2015). Semantic feature analysis targeting verbs in a quadrilingual speaker with aphasia. *Aphasiology, 29*(12), 1473–1496. https://doi.org/10.1080/02687038.2015.1049583
*Knoph, M. I. N., Simonsen, H. G., & Lind, M. (2017). Cross-linguistic transfer effects of verb-production therapy in two cases of multilingual aphasia. *Aphasiology, 31*(12), 1482–1509. https://doi.org/10.1080/02687038.2017.1358447
*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
Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149–174. https://doi.org/10.1006/jmla.1994.1008
Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition, 13*(3), 373–381. https://doi.org/10.1017/S136672891000009X
Kurland, J., Liu, A., & Stokes, P. (2018). Effects of a tablet-based home practice program with telepractice on treatment outcomes in chronic aphasia. *Journal of Speech, Language, and Hearing Research, 61*(5), 1140–1156. https://doi.org/10.1044/2018_JSLHR-L−17−0277
Kuzmina, E., Goral, M., Norvik, M., & Weekes, B. (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
*Lerman, A., Edmonds, L. A., & Goral, M. (2019). Cross-language generalisation in bilingual aphasia: What are we missing when we do not analyse discourse? *Aphasiology, 33*(9), 1154–1162. https://doi.org/10.1080/02687038.2018.1538493
*Lerman, A., Goral, M., Edmonds, L., & Obler, L. K. (2022). Strengthening the semantic verb network in multilingual people with aphasia: Within- and cross-language treatment effects. *Bilingualism: Language and Cognition*, 1–15. https://doi.org/10.1017/S1366728921001036
Lerman, A., Goral, M., & Obler, L. K. (2020). The complex relationship between pre-stroke and post-stroke language abilities in multilingual aphasia. *Aphasiology, 34* (11), 1319–1340. https://doi.org/10.1080/02687038.2019.1673303
Marte, M. J., Peñaloza, C., & Kiran, K. (2023). The cognate facilitation effect on lexical access in bilingual aphasia: Evidence from the Boston Naming Test. *Bilingualism: Language and Cognition*, 1–17. https://doi.org/10.1017/S1366728923000251
Menahemi-Falkov, M., Breitenstein, C., Pierce, J. E., Hill, A. J., O'Halloran, R., & Rose, M. L. (2021). A systematic review of maintenance following intensive therapy programs in chronic post-stroke aphasia: Importance of individual response analysis. *Disability and Rehabilitation*. https://doi.org/10.1080/09638288.2021.1955303
Norvik, M. I., Lind, M., & Jensen, B. U. (2022). Working with multilingual aphasia: attitudes and practices among speech and language pathologists in Norway. *International Multilingual Research Journal*. https://doi.org/10.1080/19313152.2021.2015935
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
Papathanasiou, I. & Coppens, P. (2021). *Aphasia and related neurogenic communication disorders* (3rd ed.). Jones & Bartlett Learning.
Papathanasiou, I., Coppens, P., & Davidson, B. (2021). Aphasia and related neurogenic communication disorders: basic concepts, management, and use of technology. In I. Papathanasiou and P. Coppens (eds.) *Aphasia and related neurogenic communication disorders* (3rd ed., pp. 3–14). Jones & Barlett Learning.
Paradis, M. (1983). Readings in aphasia in bilinguals and polyglots. *Didier*.
Paradis, M. (1993). *Foundations of aphasia rehabilitation*. Pergamon.
Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. John Benjamins.
Paradis, M., & Libben, G. (1987). *The assessment of bilingual aphasia*. Lawrence Erlbaum Associates.
*Peñaloza, C., Barrett, K., & Kiran, S. (2020). The influence of prestroke proficiency on poststroke lexical-semantic performance in bilingual aphasia. *Aphasiology, 34*(10), 1223–1240. https://doi.org/10.1080/02687038.2019.1666082
Pitres, A. (1895). Etude sur l'aphasie chez les polyglottes. Revue de Medecine, 15, 873−899. [Translated in M. Paradis (ed.), *Readings on aphasia in bilinguals and polyglots* (pp. 26−48). Marcel Didier.].
Plowman, E., Hentz, B., & Ellis, C. (2011). Post-stroke aphasia prognosis: A review of patient-related and stroke-related factors. *Journal of Evaluation in Clinical Practice, 18* (3), 689–694. https://doi.org/10.1111/j.1365−2753.2011.01650.x
R Core Team (2017) R: *A language and environment for statistical computing*. https://www.R-project.org/.
Raymer, A. M., Beeson, P., Holland, A., Kendall, D., Maher, L. M., Martin, N., Murray, L., Rose, M., Thompson, C. K., Turkstra, L., Altmann, L., Boyle, M., Conway, T., Hula, W., Kearns, K., Rapp, B., Simmons-Mackie, N., & Gonzalez Rothi, L. J. (2008). Translational research in aphasia: From neuroscience to neurorehabilitation. *Journal of Speech, Language, and Hearing Research: JSLHR, 51*(1), 259–275. https://doi.org/10.1044/1092−4388(2008/020)
Ribot, T. (1881). *Les Maladies de la Mémoire* (2nd ed.). Baillère.
Thompson, C. K. (1989). Generalization research in aphasia: A review of the literature. In T. E. Prescott (Ed.), *Clinical aphasiology* (Vol. 18, pp. 195–222). College-Hill Publication, Little, Brown and Company.
Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods, 45*(2), 576–594. https://doi.org/10.3758/s13428−012−0261−6
Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03
Webster, J., Whitworth, A., & Morris, J. (2015). Is it time to stop "fishing"? A review of generalisation following aphasia intervention. *Aphasiology, 29*(11), 1240–1264. https://doi.org/10.1080/02687038.2015.1027169

Further reading

*Abutalebi, J., Della Rosa, P. A., Tettamanti, M., Green, D. W., & Cappa, S. F. (2009). Bilingual aphasia and language control: A follow-up fMRI and intrinsic connectivity study. *Brain and Language, 109*, 141–156.
*Altman, C., Goral, M., & Levy, E. S. (2012). Integrated narrative analysis in multilingual aphasia: The relationship among narrative structure, grammaticality, and fluency. *Aphasiology, 26*, 1029–1052.
*Aziz, M. A. A., Razak, R. A., & Garraffa, M. (2020). Targeting complex orthography in the treatment of a bilingual aphasia with acquired dysgraphia: The case of a Malay/English speaker with conduction aphasia. *Behavioral Sciences, 10*(7), 109. https://doi.org/10.3390/bs10070109
*Bahadoran-Baghbaderani, A., Tahririan, M. H., Saadatnia, M., & Ketabi, S. (2021). Contributions of pre and postmorbid nondominant language interventions to coactivation of l1–l2 lexical representations: A case study of Persian-English bilingual stroke-induced aphasic patients. *Journal of Research in Applied Linguistics, 12*(1), 48–67. https://doi.org/10.22055/RALS.2021.16724
*Croft, S., Marshall, J., Pring, T., & Hardwick, M. (2011). Therapy for naming difficulties in bilingual aphasia: Which language benefits? *International Journal of Language and Communication Disorders, 46*, 48–62.
*Faroqi, Y., & Chengapppa, S. (1996). Trace deletion hypothesis and its implications for intervention with a multilingual agrammatic patient. *Osmania Papers in Linguistics, 22*(23), 79–106.
*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.
*Goral, M., Levy, E. S., & Kastl, R. (2010). Cross-language treatment generalisation: A case of trilingual aphasia. *Aphasiology, 24*, 170–187.
*Goral, M., Rosas, J., Conner, P. S., Maul, K. K., & Obler, L. K. (2012). Effects of language proficiency and language of the environment on aphasia therapy in a multilingual. *Journal of Neurolinguistics, 25*(6), 538–551. https://doi.org/10.1016/j.jneuroling.2011.06.001
*Hameau, S., & Köpke, B. (2015). Cross-language transfer for cognates in aphasia therapy with multilingual patients: A case study. *Aphasie und verwandte Gebiete, 3*, 13–19.
*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*, 169–194. https://doi.org/10.1080/02643294.2015.1061982
*Knoph, M. I. K. (2013). Language intervention in Arabic-English bilingual aphasia: A case study. *Aphasiology, 27*, 1440–1458.
*Kong, A.-P.-H., & Whiteside, J. (2015). Early recovery of a multi-lingual speaker with aphasia using Cantonese and English. *Speech, Language and Hearing, 18*, 133–139. https://doi.org/10.1179/2050572814Y.0000000059
*Kurland, J., & Falcon, M. (2011). Effects of cognate status and language of therapy during intensive semantic naming treatment in a case of severe nonfluent bilingual aphasia. *Clinical Linguistics and Phonetics, 25*(6–7), 584–600.
*Lai, W. V., Tay, R. Y. L., & Rickard Liow, S. J. (2022). Dyadic conversation training in Mandarin for bilinguals with aphasia and their communication partners. *Aphasiology, 36*(9), 1096–1126. https://doi.org/10.1080/02687038.2021.1931802
*Lee, B., Moon, H. I., Lim, S. H., Cho, H., Choi, H., & Pyun, S.-B. (2016). Recovery of language function in Korean-Japanese crossed bilingual aphasia following right basal ganglia hemorrhage. *Neurocase, 22*(3), 300–305. https://doi.org/10.1080/13554794.2016.1141966
*Lerman, A. (2022). Quantifying aphasia treatment: The effect of aphasia severity on treatment dose. *Aphasiology*. https://doi.org/10.1080/02687038.2022.2089970
*Lerman, A., Goral, M., & Obler, L. K. (2023). Rehabilitating an attrited language in a bilingual person with aphasia. *Clinical Linguistics and Phonetics, 37*(3), 272–290. https://doi.org/10.1080/02699206.2022.2052966
*Li, R., Li, W., & Kiran, S. (2021). Effect of mandarin verb network strengthening treatment (VNeST) in mandarin-english bilinguals with aphasia: A single-case experimental design. *Neuropsychological Rehabilitation, 31*, 1224–1253. https://doi.org/10.1080/09602011.2020.1773278
*Lopez, K., Gravier, M., Vasquez, A., Soto, J., Tollast, A., & Higby, E. (2022). Cross-language generalization in bilingual aphasia: influence of manner of second language acquisition. *Aphasiology*. https://doi.org/10.1080/02687038.2022.2114289
*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*, 1247–1256.
*Miertsch, B., Meisel, J., & Isel, F. (2009). Non-treated languages in aphasia therapy of polyglots benefit from improvement in the treated language. *Journal of Neurolinguistics, 22*, 135–150.
*Miller Amberber, A. (2012). Language intervention in French-English bilingual aphasia: Evidence of limited therapy transfer. *Journal of Neurolinguistics, 25*, 588–614.

*Nasrullah, R., Suganda, D., Wagiati, W., & Riyanto, S. (2019). Recovery patterns and a linguistic therapy model of Sundanese-Indonesian bilingual aphasia: A neurolinguistic study. *Indonesian Journal of Applied Linguistics, 9*(2), 452–462.

Peñaloza, C., Scimeca, M., Gaona, A., Carpenter, E., Mukadam, N., Gray, T., … 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*, Article 589330. https://doi.org/10.3389/fneur.2021.589330

*Radman, N., Spierer, L., Laganaro, M., Annoni, J. M., & Colombo, F. (2016). Language specificity of lexical-phonological therapy in bilingual aphasia: A clinical and electrophysiological study. *Neuropsychological Rehabilitation, 26*, 532–557. https://doi.org/10.1080/09602011.2015.1047383

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., Zacharewicz, M., & Gray, T. (2021). Bilingual abstract semantic associative network training (BAbSANT): A polish-english case study. *Journal of Communication Disorders, 93*, Article 106143. https://doi.org/10.1016/j.jcomdis.2021.106143

*Sazegar Nejad, S., Yadegari, F., & Teymouri, R. (2021). Persian verb and sentence production protocol (PVSP) and its efficacy on Persian-speaking agrammatic non-fluent persons with aphasia. *Aphasiology, 35*(1), 73–99.

*Veena, K., Muhammed, A., Chopra, N., Aithal, U., & Rajashekhar, B. (2012). Outcome of language-based intervention in persons with bilingual wernicke's aphasia – a case study. *Language in India, 12*, 726–744.

Watamori, T. S., & Sasanuma, S. (1978). The recovery processes of two English-Japanese bilingual aphasics. *Brain and Language, 6*(2), 127–140. https://doi.org/10.1016/0093−934X(78)90052−4