Identifying the components of a successful spoken naming therapy: A meta-analysis of word-finding interventions for adults with aphasia

Wei Ping SZE[1], Solène HAMEAU[2], Jane WARREN[1], & Wendy BEST[1]

[1] Language and Cognition, Division of Psychology and Language Sciences, University College London

[2] Department of Cognitive Science, Macquarie University, Sydney, Australia

Corresponding Author:

Wei Ping Sze,

Room 212, Language and Cognition, Division of Psychology and Language Sciences

University College London

2 Wakefield Street

London WC1N 1PF

United Kingdom

email: w.sze.17@ucl.ac.uk

Keywords: anomia rehabilitation, treatment for word retrieval difficulty, oral naming language disorder recovery, stroke, machine learning

Accepted for publication in the journal, *Aphasiology*, on 25 May 2020.

Abstract

Background:

Spoken word retrieval therapy forms an integral part of aphasia therapy. Due to the range of therapy options and variations, drawing clear-cut conclusions from the evidence base can be challenging.

Aims:

This paper consolidates recent findings (2008 - 2018), pertinent to spoken word-finding interventions. Specifically, we are interested in aphasia interventions: (1) that target single-word spoken naming; (2) whose participants are adults with acquired naming impairments after a stroke; and (3) whose treatment approach focus on the use of language forms (i.e., semantics, phonology and orthography). The over-arching objective is to determine the important therapy components underpinning successful single-word naming treatments.

Methods and Procedures:

A systematic literature search was conducted. This led to the compilation of a large-scale dataset (*n* = 222 participants from 32 papers), with the heterogenous interventions dissected into their "active" therapy components. A detailed framework ("RITA": (1) Regimen; (2) Item(s); (3) Technique(s); and (4) Application of technique(s) with their Adjuncts) was designed to organise these "active therapy ingredients". Using random forest, we identified the crucial components responsible for the successful naming of treated and untreated items respectively, at short-term and maintenance periods.

Outcomes and Results:

The role of the written form as therapeutic cues, presented either as a whole word or part-word, emerged to be a consistent and robust predictor, across the outcomes. Semantic tasks were useful in the successful naming of untreated items.

Conclusions:

Clinicians should consider using written prompts as part of spoken naming therapy. It is possible that the use of orthography underlies the success of multi-component techniques. Other clinical implications (e.g., homework, treatment intensity) are also discussed. In addition, we propose a comprehensive “RITA” framework, which summarises the “active” therapy components. “RITA” is useful for clinicians and researchers as a guide to unpack a language intervention. Furthermore, the paper highlights the strengths of a well-established method, random forest, as a valuable statistical tool to move aphasia research forward. Overall, the study refines our understanding of spoken naming treatment for those with aphasia, specifically individuals with word-finding deficits. Importantly, through the use of a robust statistical approach and an original framework designed to lay out language therapy components, the paper adds new clarity to the evidence base.

Identifying the components of a successful spoken naming therapy:

A meta-analysis of word-finding interventions for adults with aphasia

A core characteristic of aphasia is the difficulty in retrieving words (Helm-Estabrooks & Albert, 2004, p. 43) and this remains a lingering complaint of many well-recovered individuals with aphasia (Murdoch, 1990, p. 84). The principal aim of word-finding therapy is to remediate impaired spoken word retrieval. The speech and language therapist (SLT) is concerned with selecting the most effective language therapy components to achieve this outcome. Word-finding difficulty (also known as anomia or lexical-retrieval deficit) refers to the failure to retrieve the desired word, resulting in a communication breakdown (Grossman, 2014). It is characterised by semantic, phonological or neologistic paraphasias, use of non-specific words (e.g., “the thing”), circuitous language, or aborted sentences (Tesak & Code, 2008; Martin, 2013). Estimates vary, but at least one-fifth of adults hospitalised for stroke were discharged with aphasia (Ellis, Hardy, Lindrooth, & Peach, 2018).

The literature on word-finding treatments is prodigious. A PubMed search revealed that between 1973 and 1992, 70 papers were published on the subject matter. The numbers have since risen dramatically: In a similar stretch of twenty years, between 1999 to 2018, there were at least 271 new papers released – almost quadrupling the earlier quantity[1]. Understandably, for clinicians with busy schedules, this new abundance overwhelms, rather than enlightens. For each proposed treatment backed by research evidence, there are modifications of the same treatment with similarly compelling empirical provenance (e.g., Boyle and Coelho’s (1995) original Semantic Feature Analysis (SFA) was designed to treat semantic-level deficits, and in their original research, the participant had to generate six semantic features; in Hashimoto’s (2012) modified SFA, only three semantic features were required; also see Boyle, 2010, on further variations, including replacing feature generation with the “Yes/No” semantic verification tasks). To complicate matters, entirely different approaches to treat the same level of deficit, have been proposed, e.g., the use of written word-to-picture matching task to successfully treat A.E.R.'s semantic deficit (Nickels & Best, 1996b). With evidence-based practice to the fore, it is timely to ask: Which piece of evidence-based therapy method(s) should the SLTs base their practices on? Specifically, what combination of therapy components is most effective?

In this paper, we present a meta-analysis of recent literature on language-based approaches to treat spoken word retrieval deficits (2008 - 2018). The structure of the paper is as follow. First, previous reviews on word-finding therapy are summarised and evaluated. The evaluation clarifies the reasons why this meta-analysis is necessary. Next, we define and elaborate on what therapy components or "active ingredients" are. In short, they refer to specific components that make up a full-fledged language-based therapy. These encompass the more general variables (e.g., related to regimen, like the number of sessions, whether homework should be assigned) and the more treatment-specific variables, like the types of linguistic cues to provide (e.g., semantic or orthographic cues) to achieve a certain rehabilitation goal (e.g., treatment-specific effects in the short term). For coherent organisation of the multiple therapy components, we introduce the "RITA" framework, which stands for (1) Regimen; (2) Item(s); (3) Technique(s); and (4) Application of technique(s) with their Adjuncts. We proceeded to elaborate on the use of random forest as a systematic tool to navigate through this large-scale dataset (*n* = 222 participants) with the "RITA" therapy component framework as a structure. Random forest is a well-established method (elaborated below), but whose potential is yet to be fully realised in aphasiology.

The scope of this paper covers therapies that target spoken single-word production using language forms (i.e., semantics, phonology and orthography), designed for adults with word-finding difficulties after a stroke. Non-stroke populations and interventions espousing the use of alternative communication modes (e.g., alternative communication devices, Beukelman & Mirenda, 2013), total communication strategies (e.g., gestures, Rose & Douglas, 2001; drawings, Hatfield and Zangwill, 1974) or phrase and sentence productions (e.g., exchange of common phrases in constraint-induced language therapy (CILT) by Pulvermüller et al., 2001) are excluded.

Reviews and meta-analyses

Several reviews and meta-analyses of spoken word retrieval therapies for post-stroke adults have been published. In tandem with our objective to summarise literature within the past decade, the following discussion will only include the reviews and meta-analyses published from 2008 onwards. We will also explain why this review is still necessary despite these recent efforts at synthesis. Howard and Hatfield (1987), Nickels and Best (1996a), and Nickels (2002) provide qualitative overviews of earlier research. A quick summary of some recent review papers within our purview, with their more pertinent findings are presented in Table 1.[2]

[Insert Table 1 about here]

Each of these reviews attempted to address issues related to effective word-finding therapies, but from different angles and using different methodologies. A shift towards more robust objective syntheses can be discerned, with Casarin et al. (2014) being the only review relying on descriptive statistics for analysis. While collectively they add to the richness in word-finding literature, these recent reviews still do not address the issue of the relative importance of active components making up the therapies. Therapy components are the nuts and bolts of a treatment protocol. It is these ingredients that speech and language therapists are most interested in, when designing a therapy to treat a client. With the exception of de Aguiar et al. (2016), most reviews tended to either frame their discussions of therapies using broad terminologies like phonological or semantic approaches without examining the specific components; or they may simply narrow their focus and manipulate just one component (e.g., number of treatment items by Snell et al., 2010), but in doing so, forced to commit the non-realistic assumption of ceteris paribus. Consequently, outstanding issues remain unanswered.

One potential reason why several of these reviews fail to adequately address the makeup of an effective therapy could be methodological. In conventional meta-analyses, synthesis is mathematically accomplished by computing weighted effect sizes for each study, before aggregating them to derive a composite effect (Hartung, Knapp, & Sinha, 2008). In order to examine how one factor (e.g., a particular therapy component) may influence this composite effect, it is common (especially, for meta-analyses in aphasia) to run subgroup analyses as a means to isolate the effects. However, the optimal subgroup analyses (in fact, the optimal meta-analyses) take place only when the studies under inspection are relatively homogenous (see Lipsey and Wilson, 2001, pp. 158-163).[3]

Furthermore, the subgroup analyses tended to be orthogonal (i.e., variables are examined one at a time), but therapies have multiple components with a mix of categorical and continuous units of measurements. Discretising continuous measurements of independent variables will likely incur a loss of information (see Humphreys, 1978, for an analogous examination in the context of ANOVA). Importantly, an analytic approach that focuses on a component, whilst assuming the rest of them are random noise, only masks but does not address possible underlying structural differences embedded across different word-finding intervention studies. This is acknowledged by Wisenburn and Mahoney (2009), who stated that the summary statistic combines effects from both the manipulated variable *and all moderator variables*. "These moderator variables (in addition to other variables not analysed, such as participant severity, aphasia type, age, [other therapy elements] etc.) may act as a confounding variable in the interpretation of this overall score. It is functionally impossible to eliminate all potential confounding variables from an analysis of aphasia therapy" (p. 1344). Identifying the "active" therapy ingredients based on these meta-analytic findings confounded by participant and therapy heterogeneity, would be difficult, if not impossible.

In comparison, de Aguiar et al. (2016) is distinctive. It applied a robust technique (random forest) that is novel for the field, and included clinically-relevant variables for simultaneous evaluation. For all its strengths, there are two points to be made. First, despite the decision to include multiple therapy components as part of their host of variables, the search for active ingredients was not the primary focus of the authors. They cast their net wide and aimed for broad coverage: Demographic variables, various clinical diagnoses, and therapy components are all represented. Including clinical diagnoses might have rendered their analysis murkier. This is because the assessment scores used to inform the clinical diagnoses (e.g., pre-therapy status of sublexical impairment) were derived from diverse tests. The authors are frank about this: "Assessment scores … were obtained from a variety of tests, and even when tests were similar in nature they may have differed in degree of difficulty. This may have introduced noise in our dataset. Unfortunately, many studies report on results of *ad-hoc* tests, but fail to provide normative values for them (p. 12)." To eschew associated pitfalls including data noise, our paper will therefore exclude the input of clinical diagnoses as predictors.

Secondly, Auguiar et al. did not distinguish between outcomes obtained shortly post-therapy and those obtained during the longer-term maintenance phase. In their paper, these two outcomes were conflated. There are good grounds to keep the two separated. Cognitive processes involved in short term gains, may differ from those responsible for longer-term maintenance (see Cowan, 2008, and Norris, 2017, for recent reviews; also see Vitevich, Chan, & Roodenrys, 2012). Critically, within the aphasia literature, it has been demonstrated that factors (and thus, therapy components) leading to better performance immediately post-therapy, may be distinct from those resulting in superior performance when assessed much later (e.g., Howard et al., 1985; Hickin, Best, Herbert, Howard & Osbourne, 2002). In the present study, the intent was to include four outcome measures: i.e., naming performance for treated and untreated items respectively at the short-term and maintenance phases.

It is clear from this overview of recent reviews that outstanding gaps and imperfections remain. These provide motivations for a new systematic review and meta-analysis. Perhaps, the most glaring weakness is the absence of a comprehensive evaluation on the “active” therapy ingredients that underpin an intervention.

Therapy components or “active ingredients”

Naming therapies often involve multiple components. These “active” components specify the “how”, “what” and to some extent, “when” and “where” (presumably, if therapy is computerised, it could be completed from home; and “homework” as its name implies, is assumed to be done outside therapy time, and likely at home). To our knowledge, there is no published taxonomy outlining what these therapy components are. Existing checklists, like the “Template for Intervention Description and Replication” (TIDieR; Hoffmann et al., 2014), provide general guidelines on the items to report. They do not specify the individual components within an intervention.

In Tables 2 to 5, we list possible therapy components, and they are organised into four categories: Variables or information about the (1) Regimen; (2) Item(s); (3) Technique(s); and (4) Application of technique(s) with their Adjuncts (“RITA” for an acronym). This “RITA” framework aims to be comprehensive, not exhaustive, but it should be useful as a spring board for further refinement. Where possible, examples from existing literature (with priority to interventions published within the last decade) are cited in the last column of these tables. Although we apply this framework of therapy components to single-word lexical retrieval therapies, the taxonomy can be applied to language therapies beyond single-word naming.

The first set provides information on the treatment regimen (Table 2). “Regimen” variables basically describe the treatment schedule. They are the standard scheduling details clinicians convey to their clients about the expected length of the intervention (i.e., sessional information). Examples are the total number of sessions and the duration of each session etc (also see Warren, Fey, & Yodder, 2007). There is empirical evidence that these variables do influence treatment efficacy. Bhogal, Teasell, and Speechley (2003) combined results from eight studies, and found that increasing the number of hours per week (indexed by number of minutes per session and number of sessions per week in Table 2) was significantly related to improved outcomes measured on the Porch Index of Communicative Abilities. To contrast, Auguiar et al. (2016) reported that fewer sessions might in fact be beneficial, though the authors also admit they could not quite explain this paradoxical result (p. 12) – nevertheless, their findings reinforce the importance of taking into account sessional information.

[Insert Table 2 about here]

The second set of variables in Table 3 characterises information related to the items (i.e., single-words in single-word lexical retrieval treatments). Words are the raw materials used in language therapies. By extension, details on target words are vital, whether it is for therapists in communication clinics planning interventions (e.g., How many items could be included during the entire course of treatment?) or for aphasia researchers designing intervention studies.

[Insert Table 3 about here]

Each variable in the third set (Table 4) represents a technique used in therapy. A “technique” is the particular means utilised by the therapists (it is alternatively referred as “approach” by Brady et al., 2016). Variables in the third set are considered the core of the therapy. To illustrate, these are some potential “techniques” relevant for treating naming difficulties secondary to impaired phonological output processes: (a) Getting clients to generate phonological features (i.e., Phonological Component Analysis); (b) completing phonological tasks (e.g., rhyme judgment); or (c) confrontational naming assisted by part-word phonological and orthographic cues, to list a few. In clinical practice, these techniques are not mutually exclusive and can be jointly applied with proper rationales.

[Insert Table 4 about here]

The last set of variables listed in Table 5 pertains to the applications of technique(s) and adjuncts. They either elaborate on how techniques are executed (e.g., increasing or decreasing progression of cues; Baddeley & Wilson, 1994; Fillingham, Sage, & Lambon Ralph, 2005; Conroy, Sotiropoulou, Humphreys, Halai, & Lambon Ralph, 2018) or lay down general principles about conducting (e.g., the termination criteria that specifies when to cease; also see Howard, Best, & Nickels, pp. 533-534). Adjuncts that could complement therapy, like homework, are also included here.

[Insert Table 5 about here]

Random forest and its utility in word-finding therapy research

The multiple components making up the therapies present analysts with a challenge: How can the individual effects of these parts be meaningfully disambiguated? Conventional ways to conduct meta-analysis fall short. They target to calculate a pooled effect size, in order to quantitatively summarise findings obtained from a body of empirical research. Consequently, to quote, "the tidiest meta-analysis result in this regard is a distribution of effect sizes that are relatively homogenous under the fixed effects model … [However] When the distribution of effect sizes is heterogenous … the average over contrary results is not likely to converge on the truth, just muddle it." (Lipsey & Wilson, 2001, p. 162; also see Hall & Rosenthal, 1995, for criticisms on traditional approaches to meta-analysis when studies are heterogenous and complex). In the context of aphasia intervention (to which word-finding intervention belongs), discussion of heterogeneity is relevant, due to the diversity in implementations as well as patient characteristics (e.g., *SD* for effect size *d* of each therapy approach was “extremely high” (p. 1347) and ranged from “1.23” to “4.12” (p. 1344) as acknowledged by Wisenburn and Mahoney, 2009; the same brain lesion may also result in different cognitive/language impairments)[4]. Clearly, what is needed is an alternative means of analysing the wealth of word-finding intervention data without falling victim to the “pretence” of homogeneity.

Random forest is an established machine learning statistical technique that leverages the power of multiple decision trees to identify robust predictors. A full account of random forest is not within the scope of this paper. Interested readers are best referred to the original work by Breiman (2001) and his accompanying website[5] for the finer details. A rudimentary exposition to facilitate subsequent understanding is however, provided: From an original dataset, bootstrap samples (or training sets, using Breiman’s terminology) are randomly drawn. For each bootstrap sample, a unique decision tree is constructed using the principles of “Classification and Regression Trees (CART)” (Morgan & Sonquist, 1963; Breiman, Friedman, Olshen & Stone, 1984), with the nodes split by the best predictor among a subset of variables randomly chosen for that node. Multiple decision trees, each grown based on these rules, assemble to form a random forest.

Random forest is deemed suitable for our purposes because: Firstly, it has the versatility to handle data from different designs, including single case studies (see de Aguiar et al., 2016, for an example). On the other hand, classical meta-analytic approaches do not lend themselves easily to single case research: Their typical formulae set out to summarise effect sizes extracted from group studies, not single case studies (specifically, they required group mean differences and standard deviations at the group level to be entered as input, see Borenstein, Hedges, Higgins, & Rothstein, 2009; also see Shadish, Rindskopf & Hedges, 2008, for a detailed discussion on related challenges).

Secondly, a main strength of random forest is its ability to manage “wide data”, i.e., there is a large number of variables but relatively few observations (i.e., “small *n*, large *p*” or "high dimension, low sample size" problems; see Bureau et al., 2005, for a demonstration). As our meta-analysis (elaborated later) relies on a modest sample of observations (maximum *n* = 184) while seeking to examine the roles of 31 variables, random forest appears particularly suitable.

In addition, the variables examined here happen to be a mixture of categorical and continuous scales of measurement. Unlike regression (meta-regression being the case in point), this does not pose a problem for random forest, which is sufficiently robust to jointly analyse these variables with assorted scales of measurement together.

Finally, random forest produces an output quantifying the relative importance of the variable (i.e., "variable importance"). This allows us to gainfully isolate and distinguish key variables (e.g., a particular therapy component) responsible for specific treatment outcomes (e.g., perhaps therapies using phonological cues lead to successful lexical retrieval of treated items within three weeks post-therapy, but not after three weeks). Notably, variable importance of a predictor is innovatively computed by measuring the loss in predictive accuracy if its original values are substituted through permutation (Brieman, 2001; also see Hothorn, Hornik, Strobl, & Zeileis, 2019, p. 33). Conceptually, this means that during evaluation the predictor is always weighed against the entire host of variables included in the forest (i.e., the rest of the variables will have their individual predictive utilities accounted for, and they are not simply treated as random noise). This is an asset in situations when appraisal of multiple variables is required. The superiority of random forests over other statistical approaches to locate predictors is well-documented across disciplines: For example, Tagliamonte and Baayen (2012) showed how random forest could "overcome the limitations of mixed-effects models" (p. 165) using sociolinguistics data (also see Matsuki, Kuperman and Van Dyke (2016) on random forest's superiority to regression methods across three reading measures).

Purpose of this systematic review and meta-analysis

To recap, there is a lack of research that formally breaks single-word spoken naming therapy down into its multiple components for scrutiny. A novel analysis that integrates existing research findings to evaluate these various therapy components is urgently needed. The overall objective of the first part of this study is therefore to review and synthesise recent evidence obtained for language-based approaches (i.e., semantic, phonological and orthographic techniques). This collection of empirical evidence will be based on interventions designed to facilitate single-word retrieval in spoken word production for adults with impaired naming after a stroke. We include papers published between January 2008 and May 2018. This will allow us to provide an update based on the information generated roughly within a decade.

In particular, we are interested in determining the major therapy component(s) that underpin treatment efficacy, and whether the component(s) differ for treated and untreated items when assessed shortly or at longer-term periods after the therapy had terminated. We operationalised short-term as being within three weeks post-therapy, and maintenance phase as being four weeks and more post-therapy. While there are no “official” benchmarks for these time frames, our operationalisations are in fact consistent with how other studies have defined these periods (e.g., Croft, Marshall, Pring and Hardwick, 2011). Finally, random forest, given its flexibility, strengths and capacity to penetrate assorted data, is harnessed to administer the meta-analysis and uncover the “active” therapy components.

Method

Information sources and search

A systematic search was conducted using a combination of electronic databases, trial registers, and grey literature. In total, 17 sources were involved (complete list published in Appendix 1).

The keywords used in the search included stroke/cerebrovascular accident AND aphasia/naming deficit/word finding difficulty/anomia OR therapy/intervention/cue/facilitation

AND adult. Appropriate modifications were made according to the search engine's requirements. Finally, our search includes papers published from January 2008 to May 2018. Based on the procedure described, a total of 5,650 entries were generated (Figure 1).

Study selection and eligibility criteria

After removing duplicates, the first and second authors independently reviewed the 3,957 entries against the following pre-specified criteria. The studies must be single-word naming therapy research published in peer-reviewed journals or non-published theses as long as they were written in English and contained original data (i.e., research protocols without any published data are excluded, unless their data are included in the trial registers). Naming therapies conducted in non-alphabetic languages were excluded because differences in language properties could potentially necessitate different treatment requirements during naming therapies (e.g., the prevalence of homophony in say, Mandarin, may possibly limit the utility of phonological cues). We only included studies using therapy techniques targeting single-word naming, and involving language-based facilitation by way of semantic, phonological, and/or orthographic approaches. Studies excluded investigate the use of constraint-induced therapy, gestural cues, or tactile cues.

In terms of experimental designs, we included group designs (these may be randomised control trials or not), and single case experimental designs (i.e., case series and single case studies). Single case experimental designs were specifically included because in aphasia research, they constitute a substantial portion of the published evidence[6]. In terms of outcome variables, we included only studies whose single-word accuracy outcomes were obtained using the spoken picture naming paradigm. Studies with the following outcomes were thus excluded: (a) Outcomes that do not involve accuracy and/or naming (e.g., quality of life ratings); (b) accuracy outcomes that are non-verbal (e.g., gestural or written output); and (c) accuracy outcomes that are measured at the phrase or sentence level.

In terms of participants, we included studies that recruited adults with aphasia after stroke, specifically those with spoken naming difficulties. We excluded data collected from adults whose language impairment did not concern naming (e.g., impaired comprehension). We also excluded adults whose naming difficulties stem from other acquired deficits (e.g., traumatic brain injury) or neurodegenerative disorders (e.g., primary progressive aphasia).

Data extraction for meta-analysis

A total of 57 studies were identified and jointly agreed by two investigators independently as fulfilling the criteria stated. The screening process was first completed using abstracts and later, the full articles. Discrepancies between the two investigators were settled through online face-to-face discussions. Interrater reliability between the investigators was strong (κ = 0.80). Out of the 57 studies, 32 papers (Appendix 2) reported therapy-specific information at the individual participant level, and are therefore eligible for data extraction. From these papers, a dataset containing 222 participant cases was compiled. An external clinician was solicited to independently verify around 10% of the extracted data (randomly selected); no discrepancies was found.

The dataset comprised of 31 independent variables spread across five categories. Each variable was extracted at the participant-level from individual papers.

Outcome measures. Separate outcomes were extracted for treated and untreated items respectively at the short-term and maintenance phases, generating four outcome measures. To recap, short-term was operationalised as being within three weeks post-therapy, and maintenance phase as being four weeks and more post-therapy. The outcome measures were binary coded, with "1" indicating "improved naming performance". To qualify as "improved", the result must either show a significantly better naming performance as compared to baseline using statistical analyses, or report an effect size of at least a value of "4". The latter observes the benchmarks set in Beeson and Robey (2006, p. 167), which are tailored for lexical-retrieval treatments in single-case studies. Case reports whose results were inferred solely using visual analyses, e.g., split-middle lines (Morley, 2018, pp. 99-101), were excluded.

There are unfortunately no definitive guidelines on optimal numbers necessary for a random forest classifier to operate. It has been tentatively recommended for datasets to contain at least 100 observations when using general CART programs (Steinberg, n.d.). To be conservative, we pitched a minimal threshold of 120 observations for 31 predictors. For this reason, the analysis of untreated items at maintenance was dropped, as its number of observations fell below the threshold (n = 105). This leaves three outcome measures for our meta-analysis, i.e., naming performance for treated items at the short-term and maintenance phases, as well as naming performance for untreated items at the short-term phase.

Independent variables. The five categories of independent variables are: (1) Participant-related variables, (2) therapy-related variables about the regimen, (3) therapy-related variables about the items (i.e., target words); (4) therapy-related variables about the technique(s); and (5) therapy-related variables about the application of technique(s) and their adjuncts. Participant-related variables describing the demographic and clinical information of each individual were: the participant's (a) age (in years), (b) gender (male/female), (c) years of formal education received, (d) the number of months post stroke-onset and (e) baseline naming scores according to the "Boston Naming Test" (Kaplan, Goodglass, & Weintraub, 2001). These participant-related variables are included as they represent standard control factors. "Boston Naming Test" was chosen to index baseline naming performance as it emerged to be the popular choice in the papers covered in this review (140 out of 222 possible cases (approximately 63.60%) used the "Boston Naming Test" to assess initial picture naming performance; whereas 76 cases (34.23%) used another standardised instrument like the "Object and Action Naming Battery" (OANB; Druks & Masterson, 2000)).

The rest of the predictor variables are covered in depth earlier (see “Therapy components or ‘active ingredients’”; i.e., Tables 2 to 5). It is these therapy-related variables, specifying information about the treatments per se that our paper is primarily interested in.

Of note, variables specifying participants’ language deficits identified by way of baseline assessments, were not included within our set of predictors. This is due to practical concerns over reliability and heterogeneity: Across the studies reviewed, there is no uniformity in the types of assessments selected by researchers. The assessments could be various standardised tests either used in their complete formats or abridged for expediency, or unstandardised measures with unknown psychometric properties, often devised specifically for the projects (e.g., a bespoke “definition-to-written word matching task” in Macoir, Routhier, Simard & Picard, 2012). When compiled, there is substantial heterogeneity. This hodgepodge characteristic does not make for a clean variable. We therefore excluded variables pertaining to language deficits deduced by baseline assessments as predictors in our analyses.

Meta-analysis

The participant and therapy-related variables were stochastically modelled for each of the three outcome measures separately using random forest. Random forest is a supervised machine learning method. Through constructions of multiple decision trees, information is classified. The aggregated data yielded from a forest of trees is then able to provide stable and accurate predictions (Breiman, 2001). All data was processed in R (R Core Team, 2019) based on the steps described below. The implementation and programming were informed by Breiman (1999), Strobl, Hothorn, and Zeileis (2009), Tagliamonte and Baayen (2012), and de Aguiar et al. (2016).

Procedure

First, missing values were imputed following de Aguiar et al. (2016) in three steps: (a) Variables missing 56% of their raw information (e.g., 'type of stroke (ischemia vs haemorrhage)' was missing 71.62 % of its data) were excluded (rationale explained below). (b) The rfImpute function in "randomForest" was then used to impute the remaining missing values. To obtain the estimates, 100 iterations of 2000 trees were run. The missing values were imputed and adjusted using the proximity matrix generated by random forest (see Liaw & Wiener, 2018). (c) Finally, these steps were repeated 20 times to generate 20 different databases. These numbers follow figures cited in literature: For instance, Shah, Barlett, Carpenter, Nicholas, and Hemingway (2014) conducted imputation simulations on variables with a maximum of 56.8% missingness, and recommended running more than 10 repetitions for data with comparable missingness (p. 772). Similarly, de Aguair et al. were confident and ascertained their results to be reliable after running 20 repetitions (but see Wei et al., 2010, who was able to address model stability by conducting only four random forest simulations per dataset). To be conservative and in keeping with de Aguair, we simulated 20 databases to facilitate subsequent 'variable importance' computations.

Next, the "cforest function" was used to construct a random forest for each database (Hothorn, Hornik, Strobl, & Zeileis, 2019). "cforest" is an enhanced version of random forest, and corrects for multicollinearity between variables to produce unbiased variable selection for each decision tree (see Strobl, Boulesteix, Zeileis & Hothorn, 2007, who ran simulations to show that their usage of conditional permutation importance within "cforest" was able to protect against bias arising from multicollinearity). To optimise performance, we also followed the creators' recommendations and set the number of predictors randomly sampled at each node to be "5" (p. 3, Strobl, Hothorn, & Zeileis, 2009) and the number of trees grown to be "5000". In addition, to assess classification accuracy, the concordance statistic was computed using "treeresponse" command as per specified in Tagliamonte and Baayen (2012), and the "Out Of Bag" (OOB) error for each random forest extracted using "caret" (Kuhn, 2008, 2018). The concordance index provides a gross indication of the predictive ability of the classification generated by the random forest model. Based on extant literature, a concordance index of "0.5" indicates "random performance", but a value exceeding "0.83" a "strong classification model" (Hermansen, 2008). OOB error, on the other hand, is a statistic innate to random forest. It measures classification accuracy through a built-in model-testing against a unique holdout sample (subsample) not used in the training of that decision tree. A comparable OOB estimate is a maximum threshold of "0.18", based on de Aguiar et al. (2016).

In the final step, the random forest most representative of the overall findings was identified. We accomplished this in two steps for each outcome measure: (a) For every database, the variables were ranked according to their respective values of variable importance. An overall median rank was computed across the twenty databases for each variable. (b) The rankings of each database were then compared against the overall median ranks using Kendall's *Tau-b*. The latter is appropriate as it examines the compatibility of two sets of rankings by taking into account the number of paired observations that are concordant and discordant (Kendall, 1938; pp. 188-189 in Agresti, 2010)[7]. The database with the highest Kendall's rank coefficient was thus chosen to be the representative random forest for final interpretation.

Results

Based on the 222 eligible cases, results for the three outcome measures were computed, and they are reported individually below. The overall descriptive statistics for missing data are reported in percentages as follow: There were 45.33% values missing for "number of minutes per session", 39.11% for "total number of times each item was named across sessions", 35.56% for "number of sessions per week", 33.78% for "Boston Naming Test score at baseline", 28.89% for "grammatical class of treated words", 25.33% for "years of education received", 22.22% for "provision of the target word in written form", 14.67% for "mode in which therapy was delivered", 12% for "feedback per naming response", and

10.22% for the “total number of sessions”. The remaining predictors had less than 3.5% values missing; there were no missing values for 14 variables.

Successful treatment outcome of treated items post-therapy in the short-term phase

Within the 222 possible cases, 184 cases reported data for the individual’s naming performance on treated items within three weeks after therapy; out of which, 167 instances documented improvement and 17 showed no difference in their naming performance pre- and post-therapy (see Table 6 for descriptive statistics; note the descriptive statistics in this section are based on the respective representative random forest *after* the missing values had been imputed). The representative random forest produced a concordance statistic [C = 0.968] and an OOB error rate of 0.092. Both diagnostic figures fall well within the aforementioned bandwidth of good performance. These suggest that the imputed random forest classification is a strong predictive model. Kendell’s *Tau-b* coefficient showed that the rankings correlated highly and significantly with the median rankings obtained across the 20 databases [r_T = .984, p < .0001]. Figure 2 summarises our main findings based on the representative database for this measure, with the variables arranged in order of their importance. In descending order, the top three variables associated with improvement for naming treated items in the short term after therapy were: (1) Provision of the target word in the written form; (2) Explicit provision of orthographic cue(s), i.e., part-word cues like initial letter and syllables; and (3) Application of cues: Whether the application is contingent on response or compulsory regardless of response.

[Insert Table 6 and Figure 2 about here]

Successful treatment outcome of treated items post-therapy during the maintenance phase

In total, 127 identifiable cases provided individual-level naming data on treated items in a period beyond three weeks after therapy; and 116 of these reported improvements (when compared to baseline), whilst 11 detected no difference (see Table 7). The representative random forest also fulfilled the diagnostic requirements of good performance [concordance index C = 0.962; OOB = 0.087]. High and significant correlation was also found between the model's rankings and the overall median rankings [r_T = .941, p < .0001]. The main findings related to variable importance are reported in Figure 3, with (1) Feedback on accuracy of naming response, (2) Provision of the target word in the written form, and (3) Explicit provision of orthographic cue(s), found to be the three best predictors.

[Insert Table 7 and Figure 3 about here]

Successful treatment outcome of untreated items post-therapy in the short-term phase

There were 152 cases with published data for the individual's naming of the untreated items three weeks after therapy. Out of these, 27 reported an improvement, while 125 saw no difference before and after therapy (see Table 8 for descriptive statistics). The representative random forest yielded the following diagnostic statistics [C = 0.976; OOB estimate = 0.086], which indicated that the selected random forest classification is a also competent predictive model. The correlation between the representative model's rankings and the overall median rankings is strong [r_T = .987, p < .0001]. The rankings according to 'variable importance' obtained for the representative random forest model are displayed in Figure 4. The variables (1) Number of sessions for the entire therapy, (2) Use of semantic task(s), and (3) Mode in which the actual therapy was delivered (e.g., delivered on a computer or by a trained individual), emerged to be the top three predictors. Provision of target word in the written form was also amongst the top ten predictors.

[Insert Table 8 and Figure 4 about here]

Discussion

The multiple components making up a word-finding therapy are like the coloured tiles on a Rubik's cube. When we twist the cube one way to coordinate colours on the top surface, all types of unsuspecting transformations occur on the five other sides concealed from a fixed vantage point. This is why to crack the therapy riddle, there needs to be consideration of components on all sides. In this study, the multi-components in single-word spoken lexical-retrieval therapy were systematically reviewed and simultaneously accounted for using random forest analyses. We proposed the novel "RITA" framework and this guided the comprehensive coding of therapy components. In the process, evidence from 222 individual anomia cases were extracted from 32 studies. The final output were three sets of findings, one for each of the outcome measures, namely the improvement for treated items at the (1) short-term interim and (2) maintenance phase, as well as for (3) untreated items at the short-term interim.

In this paper, we did not include clinical assessment data as the predictors. As previously mentioned, this is mainly hampered by the variety of tests employed in naming therapy studies, including bespoke assessments tailored to meet particular research needs (e.g., the unpublished 'Easy Naming Set' used by Mason, Nickels, McDonald, Moses, Makin, & Taylor, 2011). There are ongoing efforts to establish a core set of standardised tests to be used in aphasia research (Wallace et al., 2019). If consensus is reached, this move towards a more uniform battery of assessments could be instrumental to drive future meta-analytic attempts forward.

Using published results as data points may result in inadvertent skewness due to publication bias. For instance, we report more cases of improvement than no improvement for the treated items (Tables 6 and 7). However, two points are worth mentioning: Firstly, there is greater awareness by editors/reviewers to appreciate null results (see "The importance of no evidence" [Editorial], 2019). In time, issues of positive publication bias would then be addressed. Secondly, the reverse is found for untreated items, where more cases of no improvement are published (Table 8). This serves to illustrate how the landscape of publication bias in aphasiology is slightly more complex, when the same papers are providing the data points for treated/untreated items. This dichotomy therefore leaves open the possibility that what has been reported is indeed a genuine reflection.

Our primary interest was to examine whether there were major therapy components that underlie treatment efficacy. There is a consistent theme across the three sets of findings: The role of orthography. Provision of the target word in the written form as well as the explicit provision of orthographic cue(s) were revealed to be particularly felicitous in the rehabilitation of treated items, be it in the short or longer term, as well as for untreated items in the short-term (Figures 2 to 4). It is also worth noting from the figures that the presentation of the written form itself is sufficient, without the ancillary need to copy the words out. Unfortunately, despite the widespread use of orthographic cues in the clinic, the orthographic approach per se is not well-explored within research[8], with greater emphasis placed on semantic approaches like SFAs (e.g., in Kiran & Thompson, 2003; van Hees et al., 2013) or phonological cueing and repetition (e.g., Biedermann & Nickels, 2008; Mason et al., 2011). In fact, in this review, none of the interventions published between January 2008 and May 2018, exclusively investigated the role of orthography. In saying so, the written form, had been frequently employed (and perhaps its clinical utility disguised) as part of a therapy using a mixture of cues (e.g., in Harnish et al., 2014, orthographic cue was used in step 2, but semantic and phonological cues were applied in steps 5 and 6), or perhaps relegated to a supplementary role in the wider context of say, assisting verbal repetition during multimodal cueing (e.g., step 1 in Menke et al., 2009) or repetition in the presence of pictures (e.g., Croot et al., 2015[9]) .

Further research will be necessary to examine why orthography works and how it operates within naming therapy (see Lorenz and Nickels, 2007, though). Three possible reasons on why the written form works in naming therapy are: (1) Visually, the print form is permanent and does not decay temporally unlike the spoken format; (2) It fosters silent reading and thus phonological recoding; and (3) It triggers the motor memory involved in writing, which in turn provides an additional pathway towards successful word retrieval. Although these accounts are speculative, they deserve investigating. Speech and language therapists are thus reminded of the clinical benefits the written form may bring, and to maximise the use of orthography to move therapy forward whenever possible (e.g., the client was literate premorbidly). Secondly, we hope the finding will inspire aphasia researchers to rethink about the role of orthography, and design relevant intervention studies to directly interrogate why it works and how to enhance its usage.

One useful finding to consider in clinical practice is that the participant's baseline naming performance did not emerge to be a strong predictor across the three outcome measures. Baseline naming performance here is indicated by the participant's Boston Naming Test score. This is potentially good news, because it suggests that participants across different levels of initial naming abilities can benefit from the various naming therapies included in the meta-analysis. In saying so, it is also possible that other participant factors are more important, like the individual's motivation or hardiness (Biel, Nitta, & Jackson, 2018), or perhaps the individual's cognitive reserve, like attention (Villard & Kiran, 2017), or short-term memory/working memory (Wright & Fergadiotis, 2012). If a set of cognitive assessments were made standardised for aphasia research, it might be possible to include these factors for a meta-analysis in the future.

Another interesting observation relates to the clinical practice of giving clients homework. In all three outcome measures, the component 'homework' did not show substantial value. It thus appeared that in the papers synthesised, giving the participants homework did not actually have an impact on their eventual naming performance. This goes against the common clinical practice of setting clients extra self-practice outside therapy hours. The reasons why homework produced weak amount of "variable importance" are not clear. It could indeed be that input during therapy suffices. Alternatively, perhaps having aphasic individuals with anomia complete homework without appropriate direct support or supervision by trained personnel, does not add therapeutic value. An intervention study manipulating the administration, i.e., completing homework with and without online support versus the absence of homework, could help tease apart these accounts.

Many of the findings from this meta-analysis also corroborate what had been previously found. For instance, published evidence suggests that a language approach whose focus is on word forms (e.g., orthography or phonology) tended to result in more treatment-specific effects, whereas an approach that targets the semantics might result in generalization to untreated items (e.g., Nickels and Best, 1996a). This trend was replicated in our findings, with the explicit provision of orthographic and phonological cues appearing in the top ten variables predicting successes for treated items in the short-term and maintenance phases (Figures 2 and 3; Tables 6 and 7), but only the orthographic format, specifically the provision of the entire target word in its written form, appearing within the top ten rankings for untreated items in the short-term (Figure 4). A different pattern was found for the use of say, semantic tasks. On the list of variables pertinent for untreated items, “semantic task” secured the second highest imputed value for variable importance (Figure 4), but it did not appear to be a predictor for treated items. In saying so, the lack of generalization effects for phonological and orthographic approaches might also reflect a publication or research bias. This is because aphasia researchers tended not to control for orthographic or phonological similarities between treated and untreated items. Instead, one suspects that in most published research, greater shared semantic overlaps might be present between treated and untreated sets of items, although this maybe unintentional (e.g., see the stimuli lists in Wright, Marshall, Wilson, and Page, 2008). Still, there is tentative evidence of generalization at the word form level (see Biederman and Nickels, 2008), but this warrants further investigation.

Another main finding replicated in this meta-analysis, is the importance of intensity, expressed in the form of the total number of teaching episodes for an item. The ‘total number of times an item was named across sessions’, for instance, appeared to influence the naming outcome of treated and untreated items in the short term (Figures 2 and 4). This is also reported elsewhere (e.g., Harnish et al., 2014, who showed that increasing the total number of teaching episodes could improve naming for all their participants).

A key contribution of this paper is its demonstration of how data from small-*n* case studies and experimental designs can be pooled together to productively explore a meaningful clinical question. In aphasia research, evidence from small-*n* intervention studies are sometimes dismissed to make way for findings based on larger-scale trials (e.g., the exclusion of single case experimental data by Brady et al., 2016). This inevitably means a substantial loss of valuable clinical information. A defence of small-*n* designs in clinical research is beyond the scope of this paper, but there are good reasons to place premium on both findings from small-*n* design studies and larger-scale clinical research, not in the least that therapists still find case studies to be useful references to gauge the existing patients on their caseloads (see Best, Sze, Edmundson, & Nickels, 2019, for a detailed discussion). While we are not the first investigators in anomia research to exploit the benefits offered by random forest, the paper adds on to the initial contribution by de Aguiar et al. (2016). Newer approaches to statistical analyses (e.g., machine learning methods) could certainly help sidestep problems conventional approaches wrestle with (e.g., the failure of a summary statistic to adequately address the inter-study differences, as raised by Wisenburn and Mahoney, 2009; also see Hall and Rosenthal, 1995), and thereby generating fresh insights.

Lastly, in outlining possible ‘active’ therapy components making up a typical naming intervention, we also propose a novel “RITA” framework: (1) Regimen; (2) Item(s); (3) Technique(s); and (4) Application of technique(s) with their Adjuncts (Tables 2 to 5). This structure is helpful for either reviewers evaluating interventions by breaking them down into their subparts or to guiding individuals to think through the nuts and bolts in therapy design. By necessity, the list of therapy components compiled is not exhaustive, but clinically purposeful. The “RITA” framework is supplied in Appendix 3, to facilitate interested users and to advance its use. The list of therapy components will always benefit from constant refinement, but the framework represents a practical start. Although designed specifically with adult word-finding therapy in mind, the “RITA” structure can easily be adapted for use in other areas within aphasia, dysarthria, and apraxia therapy, or extended to the paediatric population.

Acknowledgement

This piece of research was supported by the University College London Research Scholarships awarded to the first author.

References

Abel, S., Weiller, C., Huber, W., & Willmes, K. (2014). *Neural underpinnings for model-oriented therapy of aphasic word production*. *Neuropsychologia*, *57*, 154-165. doi: 10.1016/j.neuropsychologia.2014.03.010

Adrian, J. A., Gonzalez, M., Buiza, J. J., & Sage, K. (2011). Extending the use of Spanish Computer-assisted Anomia Rehabilitation Program (CARP-2) in people with aphasia. *Journal of Communication Disorders*, *44*(6), 666-677. doi: 10.1016/j.jcomdis.2011.06.002

Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Altmann, L. J. P., Hazamy, A. A., Carvajal, P. J., Benjamin, M., Rosenbek, J. C., & Crosson, B. (2014). Delayed stimulus-specific improvements in discourse following anomia treatment using an intentional gesture. *Journal of Speech, Language, & Hearing Research*, *57*(2), 439-454. doi: 10.1044/1092-4388(2013/12-0224)

Beeson, P. M. & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review*, *16*(4), 161-169. doi: 10.1007/s11065-006-9013-7

Best, W., Greenwood, A., Grassly, J., & Hickin, J. (2008). Bridging the gap: can impairment-based therapy for anomia have an impact at the psycho-social level? *International Journal of Language & Communication Disorders*, *43*(4), 390-407. doi: 10.1080/13682820701608001

Best, W., Sze, W. P., Edmundson, A., & Nickels, L. (2019). What counts as evidence? Swimming against the tide: Valuing both clinically informed experimentally controlled case series and randomized controlled trials in intervention research. *Evidence-Based Communication Assessment and Intervention*, *13*(3), 107-135. doi: 10.1080/17489539.2019.1597444

Beukelman, D. R., & Mirenda, P. (Eds.). (2013). *Augmentative and alternative communication: Supporting children and adults with complex communication need*s (4th ed.). Chelsea, Michigan: Paul H. Brookes Publishing Co, Inc.

Bhogal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke*, *34*(4), 987-993. doi: 10.1161/01.STR.0000062343.64383.D0

Biedermann, B. & Nickels, L. (2008). The representation of homophones: More evidence from the remediation of anomia. Cortex, 44(3), 276-293. doi: 10.1016/j.cortex.2006.07.004

Biel, M., Nitta, L., & Jackson, C. (2018). Understanding motivation in aphasia rehabilitation. In Coppens, P., & Patterson, J. (Eds.), *Aphasia rehabilitation: Clinical challenges* (pp. 393-495). Burlington, Massachusetts: Jones & Barlett Learning.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, West Sussex: John Wiley & Sons, Ltd.

Boyle, M. (2004). Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech Language Pathology*, *13*(3), 236-249. doi: 10.1044/1058-0360(2004/025)

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. doi: 10.1310/tsr1706-411

Boyle, M. (2017). Semantic treatments for word and sentence production deficits in aphasia. *Seminars in Speech and Language*, *38*(1), 52-61. doi: 10.1055/s-0036-1597256

Boyle, M. & Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech Language Pathology*, *4*(4), 94-98. doi: 10.1044/1058-0360.0404.94

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

Breiman, L. (1999). *Random forest – Random features*. (Technical Report 567). Retrieved August 1, 2018, from https://www.stat.berkeley.edu/~breiman/random-forests.pdf

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32. doi: 10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth International Group.

Bruce, C. & Howard, D. (1988). Why don't Broca's aphasics cue themselves? An investigation of phonemic cueing and tip of the tongue information. *Neuropsychologia*, *26*(2), 253-264. doi: 10.1016/0028-3932(88)90078-4

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977-990. doi: 10.3758/BRM.41.4.977

Bureau, A., Dupuis, J., Falls, K., Lunetta, K. L., Hayward, B., Keith, T. P., Van Eerdewegh, P. (2005). Identifying SNPs predictive of phenotype using random forests. *Genetic Epidemiology*, *28*(2), 171-182. doi: 10.1002/gepi.20041

Caramazza, A. & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature*, *349*, 788-789. doi: 10.1038/349788a0

Casarin, F. S., Branco, L., Pereira, N., Kochhann, R., Gindri, G., & Fonseca, R. P. (2014). Rehabilitation of lexical and semantic communicative impairments: An overview of available approaches. *Dementia & Neuropsychologia, 8*(3), 266-277. doi: 10.1590/S1980-57642014DN83000011

Coelho, C. A., McHugh, R. E., & Boyle, M. (2000). Semantic feature analysis as a treatment for aphasic dysnomia: A replication. *Aphasiology, 14*(2), 133-142. doi: 10.1080/026870300401513

Conroy, P., Sage, K., & Lambon Ralph, M. A. (2009a). A comparison of word versus sentence cues as therapy for verb naming in aphasia. *Aphasiology*, *23*(4), 462-482. doi: 10.1080/02687030802514920

Conroy, P., Sage, K., & Lambon Ralph, M. A. (2009b). Errorless and errorful therapy for verb and noun naming in aphasia. *Aphasiology*, *23*(11), 1311-1337. doi: 10.1080/02687030902756439

Conroy, P., Sage, K., & Lambon Ralph, M. (2009c). Improved vocabulary production after naming therapy in aphasia: Can gains in picture naming generalise to connected speech? *International Journal of Language & Communication Disorders*, *44*(6), 1036-1062. doi: 10.3109/13682820802585975

Conroy, P., Sage, K., & Lambon Ralph, M. A. (2009d). The effects of decreasing and increasing cue therapy on improving naming speed and accuracy for verbs and nouns in aphasia. *Aphasiology*, *23*(6), 707-730. doi: 10.1080/02687030802165574

Conroy, P. & Scowcroft, J. (2012). Decreasing cues for a dynamic list of noun and verb naming targets: A case-series aphasia therapy study. *Neuropsychological Rehabilitation*, *22*(2), 295-318. doi: 10.1080/09602011.2011.641434

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? In Progress in Brain Research, 169. doi: 10.1016/S0079-6123(07)00020-9

Croft, S., Marshall, J., Pring, T. and Hardwick, M. (2011). Therapy for naming difficulties in bilingual aphasia: Which language benefits? *International Journal of Language and Communication Disorders*, *46*(1), 48-62. doi: 10.3109/13682822.2010.484845

Croot, K., Taylor, C., Abel, S., Jones, K., Krein, L., Hameister, I., Ruggero, L., & Nickels, L. (2015). Measuring gains in connected speech following treatment for word retrieval: A study with two participants with primary progressive aphasia. *Aphasiology*, *29*(11), 1265-1288. doi: 10.1080/02687038.2014.975181

de Aguiar, V., Bastiaanse, R., & Miceli, G. (2016). Improving production of treated and untreated verbs in aphasia: A meta-analysis. *Frontiers in Human Neuroscience*, *10*(468). doi: 10.3389/fnhum.2016.00468

Dechene, L., Tousignant, M., Boissy, P., Macoir, J., Heroux, S., Hamel, M., Briere, S., & Page, C. (2011). Simulated in-home teletreatment for anomia. *International Journal of Telerehabilitation*, *3*(2), 3-10. doi: 10.5195/IJT.2011.6075

DeLong, C., Nessler, C., Wright, S., & Wambaugh, J. (2015). Semantic feature analysis: Further examination of outcomes. *American Journal of Speech-Language Pathology*, *24*(4), S864-879. doi: 10.1044/2015_AJSLP-14-0155

Di Carlo, L. M. (1980). Language recovery in aphasia: Effect of systematic filmed programmed instruction. *Archives of Physical Medicine and Rehabilitation*, *61*(1):41-44.

Druks, J., & Masterson, J. (2000). *An Object and Action Naming Battery*. Hove, East Sussex: Psychology Press.

Ellis, C., Hardy, R. Y., Lindrooth, R. C., & Peach, R. K. (2018). Rate of aphasia among stroke patients discharged from hospitals in the United States. *Aphasiology, 32*(9), 1075-1086. doi: 10.1080/02687038.2017.1385052

Grossman, M. (2014). Anomia. In M. J. Aminoff, R. B. Daroff (Eds.). (2014). *Encyclopedia of the neurological sciences: Volume 1* (2nd ed.) (pp. 194-197). San Diego, CA: Academic Press.

Hall, J. A. & Rosenthal, R. (1995). Interpreting and evaluating meta-analysis. *Evaluation & The Health Professions*, *18*(4), 393-407. doi: 10.1177/016327879501800404

Harnish, S. M., Morgan, J., Lundine, J. P., Bauer, A., Singletary, F., Benjamin, M. L., Gonzalez Rothi, L. J., & Crosson, B. (2014). Dosing of a cued picture naming treatment for anomia. *American Journal of Speech-Language Pathology*, *23*(2), S285-S299. doi: 10.1044/2014_AJSLP-13-0081

Hartung, J., Knapp, G., & Sinha, B. K. (2008). *Statistical meta-analysis with applications*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Hashimoto, N. (2012). The use of semantic- and phonological-based feature approaches to treat naming deficits in aphasia. *Clinical Linguistics & Phonetics, 26*(6), 518-553. doi: 10.3109/02699206.2012.663051

Hashimoto, N. (2016). The use of one or three semantic associative primes in treating anomia in aphasia. *American Journal of Speech-Language Pathology*, *25*(4S), S665-S686. doi: 10.1044/2016_AJSLP-15-0085

Hatfield, F. M., & Zangwill, O. L. (1974). Ideation in aphasia: The picture-story method. *Neuropsychologia*, 12, 389-393.

Helm-Estabrooks, N. & Albert, M. L. (2004). *Manual of aphasia and aphasia therapy* (2nd ed.). Austin, Texas: Pro-Ed.

Hendricks, C. T., Nicholas, M. L., & Zipse, L. (2014). Effects of phonological neighbourhood on the treatment of naming in aphasia. *Aphasiology*, *28*(3), 338-358. doi: 10.1080/02687038.2013.866209

Hermansen, S. W. (2008). Evaluating predictive models: Computing and interpreting the c statistic. Retrieved February 24, 2019, from https://support.sas.com/resources/papers/proceedings/pdfs/sgf2008/143-2008.pdf

Hickin, J., Best, W., Herbert, R., Howard, D., & Osborne, F. (2002). Phonological therapy for word-finding difficulties: A re-evaluation. *Aphasiology*, *16*(10/11), 981-999. doi: 10.1080/02687030244000509

Hoffmann, T., Glasziou, P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D. G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S. E., Dixon-Woods M, McCulloch, P., Wyatt, J. C., Chan A-W., Michie S. (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, 348: g1687. doi: 10.1136/bmj.g1687

Hothorn, T., Hornik, K., Strobl, C., & Zeileis, A. (2019). Package ‘party’: A laboratory for recursive partytioning. R package version 1.3-3.

Howard, D. & Hatfield, F. M. (1987). *Aphasia therapy: Historical and contemporary issues*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Humphreys, L. G. (1978). Doing research the hard way: Substituting analysis of variance for a problem in correlation analysis. *Journal of Educational Psychology*, *70*(6), 873-876. doi: 10.1037/0022-0663.70.6.873

Jacquemot, C., Dupoux, E., Robotham, L., & Bachoud-Lévi, A-C. (2012). Specificity in rehabilitation of word production: A meta-analysis and a case study. *Behavioural Neurology*, *25*(2), 73-101. doi: 10.3233/BEN-2012-0358

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test* (2nd ed.). Philadelphia, Pennsylvania: Lippincott Williams & Wilkins.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika, 30*(1/2), 81-93. doi: 10.2307/2332226

Kiran, S. (2008). Typicality of inanimate category exemplars in aphasia treatment: Further evidence for semantic complexity. *Journal of Speech, Language, and Hearing Research*, *51*(6), 1550-1568. doi: 10.1044/1092-4388(2008/07-0038)

Kiran, S., & Johnson, L. (2008). Semantic complexity in treatment of naming deficits in aphasia: Evidence from well-defined categories. *American Journal of Speech-Language Pathology*, *17*(4), 389-400. doi: 10.1044/1058-0360(2008/06-0085)

Kiran, S. & Thompson, C. K. (2003). The role of semantic complexity in treatment of naming deficits: Training semantic categories in fluent aphasia by controlling exemplar typicality. *Journal of Speech, Language and Hearing Research*, *46*(4), 773-787. doi: 1092-4388/03/4604-0773

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software, 28*(5), 1-26. doi: 10.18637/jss.v028.i05

Kuhn, M. (2018). caret. R package version 6.0-81. Retrieved February 1, 2019, from https://github.com/topepo/caret/

Leonard, C., Rochon, E., & Laird, L. (2008). Treating naming impairments in aphasia: Findings from a phonological components analysis treatment. *Aphasiology*, *22*(9), 923-947. doi: 10.1080/02687030701831474

Liaw, A. & Wiener, M. (2018). Package randomForest: Breiman and Cutler's random forests for classification and regression. R package version 4.6-14.

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications, Inc.

Lorenz, A. & Nickels, L. (2007). Orthographic cueing in anomic aphasia: How does it work? *Aphasiology*, *21*(6-8), 670-686. doi: 10.1080/02687030701192182

Macoir, J., Routhier, S., Simard, A., & Picard, J. (2012). Maintenance and generalization effects of semantic and phonological treatments of anomia: A case study. *Communication Disorders Quarterly*, *33*(2), 119-128. https://doi.org/10.1177/1525740110387412

Marcotte, K., Adrover-Roig, D., Damien, B., de Préaumont, M., Généreux, S., Hubert, M., & Ansaldo, A. I. (2012). Therapy-induced neuroplasticity in chronic aphasia. Neuropsychologia, 50(8), 1776-1786. doi: 10.1016/j.neuropsychologia.2012.04.001

Martin, N. (2013). Disorders of word production. In *Aphasia and related neurogenic communication disorders*. Papathanasiou, I., Coppens, P., & Potagas, C. (Eds). Burlington, Massachusetts: Jones & Barlett Learning.

Mason, C., Nickels, L., McDonald, B., Moses, M., Makin, K. & Taylor, C. (2011). Treatment of word retrieval impairments in aphasia: Evaluation of a self-administered home programme using personally chosen words. *Aphasiology*, *25*(2), 245-268. doi: 10.1080/02687038.2010.489258

Matsuki, K., Kuperman, V., & Van Dyke, J. A. (2016). The Random Forests statistical technique: An examination of its value for the study of reading. *Scientific Studies of Reading*, *20*(1), 20-33. doi: 10.1080/10888438.2015.1107073

Menke, R., Meinzer, M., Kugel, H., Deppe, M., Baumgärtner, A., Schiffbauer, H., Thomas, M., Kramer, K., Lohmann, H., Flöel, A., Knecht, S., & Breitenstein, C. (2009). Imaging short- and long-term training success in chronic aphasia. *BMC Neuroscience*, *10*, 118. doi: 10.1186/1471-2202-10-118

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Medicine*, *6*(7): e1000097. doi: 10.1371/journal.pmed1000097

Morgan, J. N. & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association*, *58*(302), 415-434. doi: 10.2307/2283276

Morley, S. (2017). *Single case methods in clinical psychology*. Abingdon, Oxon: Routledge.

Murdoch, B. E. (1990). *Acquired speech and language disorders: A neuroanatomical and functional neurological approach*. London, UK: Chapman and Hall.

Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology*, *16*(10-11), 935-979. doi: 10.1080/02687030244000563

Nickels, L. & Best, W. (1996a). Therapy for naming disorders (Part I): Principles, puzzles and progress. *Aphasiology*, *10*(1), 21-47. doi: 10.1080/02687039608248397

Nickels, L. & Best, W. (1996b). Therapy for naming disorders (Part II): Specifics, surprises and suggestions. *Aphasiology, 10*(2), 109-136. doi: 10.1080/02687039608248401

Norris, D. (2017). Short-term memory and long-term memory are still different. *Psychological Bulletin*, *143*(9), 992-1009. doi: 10.1037/bul0000108

Off, C. A., Griffin, J. R., Spencer, K. A., & Rogers, M. A. (2016). The impact of dose on naming accuracy with persons with aphasia. *Aphasiology*, *30*(9), 983-1011. doi: 10.1080/02687038.2015.1100705

Olsen, E., Freed, D. B., & Marshall, R. C. (2012). Generalisation of personalised cueing to enhance word finding in natural settings. *Aphasiology*, *26*(5), 618-631. doi: 10.1080/02687038.2012.676851

Pulvermüller, F., Neininger, B., Elbert, T., Mohr, B., Rockstroh, B., Koebbel, P., & Taub, E. (2001). Constraint-induced therapy of chronic aphasia after stroke. *Stroke*, *32*(7), 1621-1626. doi: 10.1161/01.STR.32.7.1621

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved March 16, 2019, from http://www.R-project.org

Rider, J. D., Wright, H. H., Marshall, R. C., & Page, J. L. (2008). Using semantic feature analysis to improve contextual discourse in adults with aphasia. *American Journal of Speech-Language Pathology*, *17*(2), 161-172. doi: 1058-0360/08/1702-0161

Robbey, R. R. (1998). A meta-analysis of clinical outcomes in the treatment of aphasia. *Journal of Speech, Language, and Hearing Research*, *41*(1), 172-187.

Rose, M. & Douglas, J. (2001). The differential facilitatory effects of gesture and visualisation processes on object naming in aphasia. *Aphasiology*, *15*(10-11), 977-990. doi: 10.1080/02687040143000339

Rose, M. & Douglas, J. (2008). Treating a semantic word production deficit in aphasia with verbal and gesture methods. *Aphasiology*, *22*(1), 20-41. doi: 10.1080/02687030600742020

Routhier, S., Bier, N., & Macoir, J. (2015). The contrast between cueing and/or observation in therapy for verb retrieval in post-stroke aphasia. *Journal of Communication Disorders*, *54*, 43-55. doi: 10.1016/j.jcomdis.2015.01.003

Routhier, S., Bier, N., & Macoir, J. (2016). Smart tablet for smart self-administered treatment of verb anomia: Two single-case studies in aphasia. *Aphasiology*, *30*(2-3), 269-289. doi: 10.1080/02687038.2014.973361

Sage, K., Snell, C., & Lambon Ralph, M. A. (2011). How intensive does anomia therapy for people with aphasia need to be? *Neuropsychological Rehabilitation*, *21*(1), 26-41. doi: 10.1080/09602011.2010.528966

Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention*, *2*(3), 188-196. doi: 10.1080/17489530802581603

Shah, A. D., Barlett, J. D., Carpenter, J., Nicolas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using MICE: A CALIBER study. *American Journal of Epidemiology*, *179*(6), 764-774. doi: 10.1093/aje/kwt312

Silkes, J. P., Dierkes, K. E., & Kendall, D. L. (2013) Masked repetition priming effects on naming in aphasia: A phase I treatment study. *Aphasiology*, *27*(4), 381-397. doi: 10.1080/02687038.2012.745475

Snell, C., Sage, K., & Lambon Ralph, M. A. (2010). How many words should we provide in anomia therapy? A meta-analysis and a case series study. *Aphasiology*, *24*(9), 1064-1094. doi: 10.1080/02687030903372632

Steinberg, D. (n. d.). Rules of thumb when working with small data samples. Retrieved from https://www.salford-systems.com/blog/dan-steinberg/rules-of-thumb-when-working-with-small-data-samples

Strobl, C., Boulesteix, A-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*, 25. doi: 10.1186/1471-2105-8-25

Strobl, C., Hothorn, T., & Zeileis, A. (2009). Party on! A new, conditional variable importance measure for random forests available in the party package. *The R Journal*, *2*, 14-17.

Tagliamonte, S. A. & Baayen, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, *24*(2), 135-178. doi: 10.1017/S0954394512000129

Tesak, J. & Code, C. (2008). *Milestones in the history of aphasia: Theories and protagonists*. Hove, East Sussex: Psychology Press.

The importance of no evidence. [Editorial]. (2019, March). *Nature Human Behaviour*, *March 2019* (3), 197. doi: 10.1038/s41562-019-0569-7

van Hees, S., Angwin, A., McMahon, K., & Copland, D. (2013). A comparison of semantic feature analysis and phonological components analysis for the treatment of naming impairments in aphasia, *Neuropsychological Rehabilitation*, *23*(1), 102-132. doi: 10.1080/09602011.2012.726201

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience & Biobehavioral Reviews*, *35*(3), 407-426. doi: 10.1016/j.neubiorev.2010.04.007

Villard, S. & Kiran, S. (2017). To what extent does attention underlie language in aphasia? *Aphasiology*, *31*(10), 1226-1245. doi: 10.1080/02687038.2016.1242711

Vitali, P., Tettamanti, M., Abutalebi, J., Ansaldo, A.-I., Perani, D., Cappa, S. F., & Joanette, Y. (2010). Generalization of the effects of phonological training for anomia using structural equation modelling: A multiple single-case study. *Neurocase*, *16*(2), 93-105. doi: 10.1080/13554790903329117

Vitevitch, M. S., Chan, K. Y., Roodenrys, S. (2012). Complex network structure influences processing in long-term and short-term memory. *Journal of Memory & Language*, *67*(1), 30-44. doi: 10.1016/j.jml.2012.02.008

Wallace, S. E. & Kimelman, M. D. Z. (2013). Generalization of word retrieval following semantic feature treatment. *NeuroRehabilitation*, *32*(4), 899-913. doi: 10.3233/NRE-130914

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., Babbitt, E., Bose, A., Brady, M., Cherney, L. R., Copland, D., Cruice, M., Enderby, P., Hersh, D., Howe, T., Kelly, H., Kiran, S., Laska, A-C., Marshall, J., Nicholas, M., Patterson, J., Pearl, G., Rochon, E., Rose, M, Sage, K., Small, S. and Webster, J. (2019). A core outcome set for aphasia treatment research: The ROMA consensus statement. *International Journal of Stroke*, *14*(2), 180-185. doi: 10.1177/1747493018806200

Wambaugh, J. L., Mauszycki, S., & Wright, S. (2014). Semantic feature analysis: Application to confrontation naming of actions in aphasia. *Aphasiology*, *28*(1), 1-24. doi: 10.1080/02687038.2013.845739

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation & Developmental Disabilities Research Reviews*, *13*(1), 70-77. doi:10.1002/mrdd.20139

Webster, J. and Whitworth, A. (2012). Treating verbs in aphasia: exploring the impact of therapy at the single word and sentence levels. *International Journal of Language & Communication Disorders*, *47*(6), 619-636. doi: 10.1111/j.1460-6984.2012.00174.x

Wei, C-L., Rowe, G. T., Escobar-Briones, E., Boetius, A., Soltwedel, T., Caley, M. J., Soliman, Y., Huettmann, F., Qu, F., Yu, Z., Pitcher, C. R., Haedrich, R. L., Wicksten, M. K., Rex, M. A., Baguley, J. G., Sharma, J., Danovaro, R., MacDonald, I. R., Nunnally, C. C., Deming, J. W., Montagna, P., Levesque, M., Weslawski, J. M., Wlodarska-Kowalczuk, M., Ingole, B. S., Bett, B. J., Billett, D. S. M., Yool, A., Bluhm, B. A., Iken, K., Narayanaswamy, B. E. (2010). Global patterns and predictions of seafloor biomass using random forests. *PLoS ONE*, *5*(12), e15323. doi: 10.1371/journal.pone.0015323

Weill-Chounlamountry, A., Capelle, N., Tessier, C., & Pradat-Diehl, P. (2013). Multimodal therapy of word retrieval disorder due to phonological encoding dysfunction. *Brain Injury*, *27*(5), 620-631. doi: 10.3109/02699052.2013.767936

Wisenburn, B. & Mahoney, K. (2009). A meta-analysis of word-finding treatments for aphasia. *Aphasiology*, *23*(11), 1338-1352. doi: 10.1080/02687030902732745

Woolf, C., Caute, A., Haigh, Z., Galliers, J., Wilson, S., Kessie, A., Hirani, S., Hegarty, B., & Marshall, J. (2016). A comparison of remote therapy, face to face therapy and an attention control intervention for people with aphasia: A quasi-randomised controlled feasibility study. *Clinical Rehabilitation*, *30*(4), 359-373. doi: 10.1177/0269215515582074

Wright, H. H. & Fergadiotis, G. (2012). Conceptualising and measuring working memory and its relationship to aphasia. *Aphasiology*, *26*(3/4), 258-278. doi: 10.1080/02687038.2011.604304

Wright, H. H., Marshall, R. C., Wilson, K. B., & Page, J. L. (2008). Using a written cueing hierarchy to improve verbal naming in aphasia. *Aphasiology*, *22*(5), 522-536. doi: 10.1080/02687030701487905

Footnotes

[1] The PubMed search was performed by the first author on 08 March 2019.

[2] It is beyond the scope of this paper to include all reviews published between 2008 and 2018. This is not intended to be a systematic report of existing reviews of single-word naming therapies. Rather, Table 1 represents the main efforts, based on the following criteria. These reviews: (1) Evaluated the efficacies between different interventions, i.e., not just devoted to evaluating one type of therapy; and (2) Provided numerical analysis of the studies they included, to support their main points. It is common in aphasia research for the reviews to include single-case experimental designs and/or randomised controlled trials (RCTs).

[3] Technically, researchers could just run say subgroup analyses with just studies looking at Semantic Feature Analyses (SFAs), and obtain the pooled effect. This finding is however quite limited, as it only tells us in isolation, whether SFA is effective, but fails to help us compare against other interventions. Additionally, it treats SFA as a package, and does not recognise it has shared similarities with other types of therapies (e.g., feedback, written semantic cues). In truth, any of these shared components could be the genuine cause behind its proposed efficacy.

[4] Unfortunately, in the Cochrane review by Brady et al. (2016), the heterogeneity index, I^2 statistic, was not reported.

[5] https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

[6] A rudimentary count revealed that for the journal “Aphasiology”, single case or case series intervention studies made up 20% of the original articles published in 2018

[7] Although not stated in the analyses, for each database, we created scatterplots between the median rankings across 20 databases and the individual database’s rankings to ensure that in each of them, the relationship between the two types of rankings was monotonic. All scatterplots showed a linear distribution in the increasing direction. Monotonicity is an assumption in Kendall’s *Tau-b*.

[8] To clarify, we are referring to the lack of studies that investigate the exclusive role of orthography, i.e., provide only the written form in treatment and is not accompanied by other

cues (e.g., phonology). Orthographic cues are often explored in aphasia rehabilitation literature in the context of other techniques/cues, but it also means we cannot disambiguate the role of orthography from these accompaniments. We would also like to mention that the figures in Tables 6 to 8 do not contradict our statement on the rarity of research dedicated to the sole investigation of orthographic cues, compared to other approaches. In these tables, the numbers reflect the *number of participants*, not the number of studies. Snell et al. (2010), who provided both phonological and orthographic cues, was a case series of 13 participants. Rider et al. (2008), who used SFA, had three participants. This is not surprising because of the resource demands (e.g., time) necessitated when conducting SFA, which may explain the smaller sample sizes. Nevertheless, all these suggest that many participants were involved in research that examined written cues in the wider context of other techniques/cues. These numbers further emphasise how fewer participants (in our case, zero participants across the 32 papers) undertook a therapy research that relied purely on orthography.

[9] Croot et al. (2015) is mentioned to make an illustration on the therapy technique. In their paper, the two participants were diagnosed with primary progressive aphasia, so their data are not included in this meta-analysis.

Appendix 1

List of information sources

(1) For electronic databases, we included the more general search engines (PsychINFO, Web of Science, Embase, PubMed, Cochrane Library Databases) as well as discipline-specific search engines (for allied health: Allied and Complementary Medicine (AMED), Cumulative Index to Nursing and Allied Health Literature (CINAHL Plus); for speech and language therapy: Speech Pathology Database for Best Interventions and Treatment Efficacy (SpeechBITE)).

(2) For trial registers, we included the following: ClinicalTrials.gov, the Stroke Trials Registry (http://www.strokecenter.org/trials/), Current Controlled Trials, International Standard Registered Clinical/soCial sTudy Number (ISRCTN) registry, World Health Organization International Clinical Trials Registry Platform (WHO ICTRP), Database of Research in Stroke (DORIS) and the Cochrane Central Register of Controlled Trials.

(3) Lastly, grey literature was trawled using ProQuest for unpublished theses/dissertations and by hand-searching reference lists of other reviews for articles not identified electronically.

Appendix 2

Articles eligible for meta-analysis

SN	Author(s)	Year	Title of article	Journal
1	Abel, S., Weiller, C., Huber, W., Wilmes, K., & Specht, K.	2015	Therapy-induced brain reorganization patterns in aphasia	Brain
2	Adrian, J. A., Gonzalez, M., Buiza, J. J., & Sage, K.	2011	Extending the use of Spanish Computer-assisted Anomia Rehabilitation Program (CARP-2) in people with aphasia	Journal of Communication Disorders
3	Conroy, P., Sage, K., & Lambon Ralph, M. A.	2009	A comparison of word versus sentence cues as therapy for verb naming in aphasia	Aphasiology
4	Conroy, P., Sage, K., & Lambon Ralph, M. A.	2009	Errorless and errorful therapy for verb and noun naming in aphasia	Aphasiology
5	Conroy, P., Sage, K., & Lambon Ralph, M. A.	2009	Improved vocabulary production after naming therapy in aphasia: can gains in picture naming generalise to connected speech?	International Journal of Language and

				Communication Disorders
6	Conroy, P., Sage, K., & Lambon Ralph, M. A.	2009	The effects of decreasing and increasing cue therapy on improving naming speed and accuracy for verbs and nouns in aphasia	Aphasiology
7	Conroy, P. & Scowcroft, J.	2012	Decreasing cues for a dynamic list of noun and verb naming targets: A case-series aphasia therapy study	Neuropsychological Rehabilitation
8	Dechene, L., Tousignant, M., Boissy, P., Macoir, J., Heroux, S., Hamel, M., Briere, S., & Page, C.	2011	Simulated in-home teletreatment for anomia	International Journal of Telerehabilitation
9	DeLong, C., Nessler, C., Wright, S., & Wambaugh, J.	2015	Semantic feature analysis: Further examination of outcomes	American Journal of Speech-Language Pathology
10	Harnish, S. M., Morgan, J., Lundine, J. P., Bauer, A., Singletary, F., Benjamin, M. L., Gonzalez Rothi, L. J., & Crosson, B.	2014	Dosing of a cued picture-naming treatment for anomia	American Journal of Speech-Language Pathology

11	Hashimoto, N.	2016	The use of one or three semantic associative primes in treating anomia in aphasia	American Journal of Speech-Language Pathology
12	Hendricks, C. T., Nicholas, M. L., & Zipse, L.	2014	Effects of phonological neighbourhood on the treatment of naming in aphasia	Aphasiology
13	Jacquemot, C., Dupoux, E., Robotham, L., & Bachoud-Levi, A. C.	2012	Specificity in rehabilitation of word production: A meta-analysis and a case study	Behavioural Neurology
14	Kiran, S.	2008	Typicality of inanimate category exemplars in aphasia treatment: Further evidence for semantic complexity	Journal of Speech, Language, and Hearing Research
15	Kiran, S. & Johnson, L.	2008	Semantic complexity in treatment of naming deficits in aphasia: Evidence from well-defined categories	American Journal of Speech-Language Pathology
16	Macoir, J., Routhier, S., Simard, A., & Picard, J.	2012	Maintenance and generalization effects of semantic and phonological treatments of anomia: A case study	Communication Disorders Quarterly

17	Marcotte, K., Adrover-Roig, D., Damien, D., de Preaumont, M., Genereux, S., Hubert, M., Ansaldo, A. I.	2012	Therapy-induced neuroplasticity in chronic aphasia	Neuropsychologia
18	Menke, R., Meinzer, M., Kugel, H., Deppe, M., Baumgärtner, A., Schiffbauer, H., Thomas, M., Kramer, K., Lohmann, H., Flöel, A., Knecht, S., & Breitenstein, C.	2009	Imaging short- and long-term training success in chronic aphasia	BMC Neuroscience
19	Off, C. A., Griffin, J. R., Spencer, K. A., & Rogers, M. A.	2016	The impact of dose on naming accuracy with persons with aphasia	Aphasiology
20	Olsen, E., Freed, D. B., & Marshall, R. C.	2012	Generalisation of personalised cueing to enhance word finding in natural settings	Aphasiology
21	Rider, J. D., Wright, H. H., Marshall, R. C., & Page, J. L.	2008	Using semantic feature analysis to improve contextual discourse in adults with aphasia	American Journal of Speech-Language Pathology
22	Rose, M. & Douglas, J.	2008	Treating a semantic word production deficit in aphasia with verbal and gesture methods	Aphasiology

23	Routhier, S., Bier, N., & Macoir, J.	2015	The contrast between cueing and/or observation in therapy for verb retrieval in post-stroke aphasia	Journal of Communication Disorders
24	Routhier, S., Bier, N., & Macoir, J.	2016	Smart tablet for smart self-administered treatment of verb anomia: Two single-case studies in aphasia	Aphasiology
25	Sage, K., Snell, C., & Lambon Ralph, M. A.	2011	How intensive does anomia therapy for people with aphasia need to be?	Neuropsychological Rehabilitation
26	Snell, C., Sage, K., & Lambon Ralph, M. A.	2010	How many words should we provide in anomia therapy? A meta-analysis and a case series study.	Aphasiology
27	van Hees, S., Angwin, A., McMahon, K., & Copland, D.	2013	A comparison of semantic feature analysis and phonological components analysis for the treatment of naming impairments in aphasia	Neuropsychological Rehabilitation
28	Vitali, P., Tettamanti, M., Abutalebi, J., Ansaldo, A.-I., Perani, D., Cappa, S. F., & Joanette, Y.	2010	Generalization of the effects of phonological training for anomia using structural equation modelling: A multiple single-case study.	Neurocase
29	Wallace, S. E. & Kimelman, M. D. Z.	2013	Generalization of word retrieval following semantic feature treatment	NeuroRehabilitation

30	Wambaugh, J. L., Mauszycki, S., & Wright, S.	2014	Semantic feature analysis: Application to confrontation naming of actions in aphasia	Aphasiology
31	Weill-Chounlamountry, A., Capelle, N., Tessier, C., & Pradat-Diehl, P.	2013	Multimodal therapy of word retrieval disorder due to phonological encoding dysfunction	Brain Injury
32	Wright, H. H., Marshall, R. C., Wilson, K. B., & Page, J. L.	2008	Using a written cueing hierarchy to improve verbal naming in aphasia	Aphasiology

Appendix 3

List of therapy components based on the “RITA” framework

Information about Regimen	
No. of minutes per session	
No. of sessions per week	
No. of sessions for entire therapy	
Information about Items	
Total no. of items treated across sessions	
Total no. of times each item was named across sessions	
Grammatical class of treated words	N, V etc.
Information about Techniques	
Provision of semantic cue(s): Explicit	√: X :
Provision of phonological cue(s): Explicit	√: X :
Provision of orthographic cue(s): Explicit	√: X :
Provision of phonological cue(s): Implicit	√: X :
Provision of the target word in spoken form	√: X :
Provision of the target word in written form	√: X :
Spoken repetition of target word: Compulsory	√: X :
Spoken repetition of target word: Contingent on response	√: X :

(continued) Techniques	
Copying of written target word: Compulsory	√: X :
Semantic cue(s) generation	√: X :
Phonological cue(s) generation	√: X :
Semantic task(s)	√: X :
Orthographic task(s)	√: X :
Information abt Technique Application	
Application of cues: Progression	↑: ↓:
Application of cues: Depends on response vs. compulsory vs. mixture	
Feedback on accuracy of the naming response	Explicit: Implicit: None:
Termination criteria	
Delivery mode for therapy: Computer; in-person; mixture	
Homework	√: X :

Table 1. Selected reviews and meta-analyses (2009 to 2018).

Paper / Brief description, including aim(s)	Main finding(s)
(1) Wisenburn and Mahoney (2009): • 44 articles, based on single-case experimental designs and group studies, published from 1989 and 2007, yielded 107 effect sizes. • The aim was to compare the effect sizes (to represent treatment efficacy) for each of the main word-finding approaches (semantic, phonologic, mixed). Effect sizes were also evaluated based on moderators, like the types of words examined (treated, untreated etc.), and duration (immediately post-therapy, three-months post-therapy etc.).	Subgroup comparisons of effect sizes were computed. • Importantly, the phonological approach (a loose category that includes phonemic cues, orthographic cues, verbal repetition and writing words etc.) yielded the largest magnitude, whether it was immediately post-therapy or in the three-month follow-up. • Also, "semantic therapy appeared to generalise slightly more than either phonological or mixed therapies to the unexposed words" (p. 1345 - 1346). The authors were however cautious, stating that "no definite conclusions could be made as to whether semantic, phonologic, or mixed therapy is most efficacious" (p. 1346). This is in part because these mean effect sizes combined "data across all moderator variables" (p. 1344; e.g., participant severity, age, different components across therapies etc.), thus lacking the precision to permit firm conclusions.
(2) Snell, Sage, and Lambon Ralph (2010): • 21 articles, based on single-case experimental designs, published between 1985 and 2006, yielded 109 individual data points. • The aim was to investigate (a) whether the number of items used in treatment affected the outcome, and (b) whether the severity of the patients' word-finding deficit influenced the outcome.	Correlation analyses were run between naming outcome (computed as 'proportion gain', p. 1067) and (a) number of items used in treatment, as well as (b) severity (operationalised as scores obtained in Boston Naming Test). • The results were not conclusive. This is partly because a bias in literature was identified, in that the more severe participants were assigned more treatment items. • A follow-up case series based on 13 stroke participants suggested that participants, regardless of severity, can tolerate more items (e.g., *n* = 60) in single-word naming therapies.
(3) Jacquemot, Dupoux, Robotham, and Bachoud-Levi (2012): • 39 articles, based on single-case experimental designs, published between	Signal detection theory was used for analysis. The authors argued that sensitivity index (*d*') was a bias-free measure of rehabilitation sensitivity, and thus appropriate for their

1990 to November 2009, yielded 124 individual data points. • The aim was to identify which processing components (boxes and arrows) within a neuropsychological model are most important for treatment, in order to generate successful outcomes for spoken word-finding. This is different from the other meta-analyses, which examined treatment components, not the 'boxes and arrows' of a theoretical model.	analysis. d' was calculated for each processing component in the neuropsychological model. In this context, d' accounts for whether the treatment outcome is successful and whether the treatment task activates the intended component (e.g., Hits = Successful outcome, task correctly activates intended component; see Jacquemot et al., p. 87). • Rehabilitation was most successful if the components (a) 'phonological output lexicon' and (b) 'phonological output' were activated. • Interestingly, if pathways linking (a) 'orthographic input' and 'phonological output', as well as (b) 'orthographic input lexicon' and 'phonological output lexicon' were triggered, rehabilitation was successful too. However, not all tasks were equally beneficial for treating spoken word-finding. Tasks activating the phonological input and phonological input lexicon (e.g., syllable counting), were not beneficial.
(3) Casarin, Branco, Pereira, Kochhann, Gindri, and Fonseca (2014): • 28 articles, including those using single-case experimental designs and RCTs, published between 2005 and May 2014 were summarised. • The aim was to evaluate lexical-semantic interventions (this was defined broadly and included interventions like gestural facilitation training and CILTs) • However, the target population were not simply adults with stroke, but also those with traumatic brain injury, and/or dementia. • The outcomes of these studies also do not appear to be specific to spoken single-word naming either: E.g., improvements in functional communication measured by the Amsterdame Nijmegen Everyday Language Test (ANELT) following a broad package of cognitive-linguistic treatment (de Jong-Hagelstein et al., 2011).	Descriptive statistics was used for analyses. • The authors reported that majority of the studies (39.28% of them) investigated using semantic cues, followed by SFA (17.86%). After these semantic-based interventions, phonological interventions appeared to be next in terms of popularity (17.85%) and then a miscellany of other types of interventions (e.g., intention and pantomime gestures, Copy and Recall Treatment).
(5) Brady, Kelly, Godwin, Enderby and Campbell (2016): • 57 RCTs, involving 3002 participants, for all studies published until September 2015.	• 29 subgroup analyses of effect sizes on the outcome of 'expressive language: naming'

This, however, includes *any speech and language therapies* seeking to improve communication and participation; and these are *not specific to word-finding*. • The aim was to provide a comprehensive Cochrane review into speech and language therapy after a stroke. • The meta-analysis of RCTs examined many topical issues (e.g., therapy versus no therapy) on various outcomes (e.g., functional communication, receptive language, expressive language, economic outcomes etc.). For our purposes, only the subset examining 'expressive language: naming' is of interest.	was done (see Brady et al., pp. 'i' to 'vi', for the complete listing). • There was no evidence suggesting benefit or harm for most of the comparisons. A selection of these comparisons with "no difference found" is provided here: (a) 'Therapy received' versus 'no therapy received' (this comparison is based on 275 participants across 7 trials; though quality of evidence was rated as 'poor', p. 43); (b) 'High-intensity' versus 'low-intensity' therapy (this is based on 2 studies; intensity refers to number of weeks per hour)[1]; (c) Therapies with short versus long durations (this is based on 3 studies; duration refers to the total number of weeks)[2]; and (d) Semantically-based therapy versus CILT[3] (this is just based on 1 study, i.e., Wilssens et al., 2015). • Interestingly, participants whose therapies were mediated by a computer performed better on measures of untreated items, compared to those in direct contact with an actual therapist (based on 2 studies; p. 38 in Brady et al., 2016).
(6) de Aguiar, Bastiaanse, and Miceli (2016): • 30 articles, based on single-case experimental designs, published between 1992 to 2014, yielded 166 individual data points. • The aim was to examine the predictive values of over 30 demographic (e.g., age, gender), clinical (e.g., type of impairment: semantic, sublexical etc.) and treatment-related (semantic cues, phonemic cues, cue direction etc.) variables when they are entered collectively, for two separate outcomes specific to verbs (i.e., improvement on treated vs untreated verbs).	Analysis was completed using random forest. • For improvement on treated verbs, good verb comprehension, good repetition skills, and number of therapy sessions attended per week were found to be important predictors. • For improvement on untreated verbs, incorporation of morphological cues in therapy, presence of grammatical impairment, presence of semantic impairment (indexed by noun comprehension) and number of therapy sessions attended per week were found to be important predictors.

[1] Intensity was measured in terms of hours per week, and exact operationalization varied between studies (p. 30; Brady et al., 2016). In Brady et al., high-intensity was represented by Pulvermuller (2001) and SP-i-RiT with 10 hours per week and low-intensity being 2 hours in SP-i-RiT and 5 hours in Pulvermuller (2001).

[2] Duration was measured in terms total number of weeks. Exact operationalization again varied between studies (p. 34; Brady et al., 2016). Short duration in Brady et al. could be two weeks (Pulvermuller, 2001) or between 6 and 9 months (Di Carlo, 1980) and long duration could be between 3 and 5 weeks (Pulvermuller, 2001) or between 5 and 22 months (Di Carlo, 1980).

[3] Unfortunately, this is the only comparison made between approaches in this review that is pertinent for us (pp. 38-39), and technically, CILT should be excluded from our discussion, as the linguistic unit during therapy is beyond the single-word.

Table 2. Information or variables describing the regimen.

Variable	Operationalization	Example from Literature
Number of minutes per session Scale of measurement: Ratio	Refers to the average length of time for each session. This provides insights on the temporal space made available for the items within one therapy slot.	Woolf et al. (2016): Participants (*n* = 5) were assigned remote therapy to be done at home, using mainstream video conferencing (skype or FaceTime) with the therapist who called from the hospital clinic. There were eight sessions with two sessions per week. Each participant completed eight hours of therapy (i.e., each session was an hour). Results suggested that the individuals who completed remote therapy from the hospital clinic, improved in their naming for treated and untreated items.
Number of sessions per week Scale of measurement: Ratio	Refers to the average number of sessions per unit of time (laid down as x1 week here for uniformity). This provides insights on how regular sessions are. This is also known as “dose frequency” in Warren et al.’s (2007).	
Number of sessions for the entire therapy Scale of measurement: Ratio	Refers to the total number of sessions provided. This provides insights on how long the entire course of therapy takes.	

Table 3. Information or variables describing the items.

Variable	Operationalization	Examples from Literature
Total number of items treated across sessions Scale of measurement: Ratio	Refers to the cumulative number of items targeted during the span of therapy. This provides numerical insights on the item coverage.	Snell et al. (2010) compared naming performance for the same individuals pre- and post-naming therapy for a 60-item set versus a 20-item set. Although overall analyses in Snell et al.'s data suggested that regardless of naming deficit severity, participants could tolerate more items, individual analyses revealed that at least two individuals (SS and IH) actually performed better proportionately on smaller sets of items (p. 1087). This highlights the importance of considering the number of items intended for therapy when planning interventions.
Total number of times each item was named across sessions Scale of measurement: Ratio	Refers to the cumulative number of times each item was named during the span of therapy. Also called 'cumulative intervention intensity' in Warren et al.'s vernacular. This provides insights on the amount of naming practice provided for each item.	Harnish et al. (2014) tracked the number of naming attempts, and found that at least 2 participants in their sample (*n* = 8) required 1200 naming attempts to achieve significant gains, compared to baseline performance. In other words, the number of times an item is named matter.
Grammatical class of treated words Scale of measurement: Category – Noun, verb, mixture of nouns and verbs	Refers to whether the target words are nouns, verbs or adjectives etc. Only nouns and verbs (or a mix of these two word classes) happened to be covered by the studies included in this review.	In the initial published trials of Semantic Feature Analysis, the researchers targeted nouns (e.g., Boyle, 2004), whereas a similar feature-generation treatment like Verb Network Strengthening Treatment (VNeST) targeted verbs as implied by its namesake (Edmonds, 2014).

	The debate on whether and how words of different grammatical classes are organised or processed differently is a long-standing debate, but beyond the scope of this paper (see Caramazza & Hillis, 1991 or Vigliocco, Vinson, Druks, Barber, & Cappa, 2011, for the theoretical accounts on lexical organisation of nouns versus verbs). Similar therapies may be designed to target different word classes. They may also yield varying levels of success (see Webster and Whitworth, 2012, whose qualitative review suggested that while verbs respond to similar treatments, their gains might be more limited than nouns). Combined, this means clinicians should be mindful of the grammatical classes of the words selected for treatment, which could in turn affect other therapy components (e.g., choice of therapy technique(s) etc.).	Early findings supportive of Semantic Feature Analysis were published by the original researchers, with emphatic improvements on treated and untreated items reported (Boyle & Coelho, 1995; Coelho, McHugh, & Boyle, 2000; see Boyle, 2017, for a review and also how Semantic Feature Analysis had been later modified to accommodate verbs). Recent evidence, however, had been mixed (e.g., van Hees, Angwin, McMahon, & Copland, 2013: Four out of their eight participants did not show item-specific improvement; for five in eight participants, there was no generalisation to untreated items).

Table 4. Information or variables describing the techniques.

Variable	Operationalization	Examples from Literature
Provision of semantic cue(s): Explicit Scale of measurement: Binary – Yes, No	Whether semantic cue(s) is/are explicitly provided during the actual naming task. Examples include: Presenting the definition, describing key features of the target item and cloze sentence completion. The cues can be delivered in the spoken and/or written form(s). This variable excludes the semantic cues generated and/or provided during say, Semantic Feature Analysis (SFA). SFA is a multi-step program, distinct from the basic direct application of semantic cues (Boyle & Coelho, 1995).	Abel, Weiller, Huber, and Willmes (2014): One of the therapies investigated involved a hierarchy of semantic cues[4]. The order of semantic cues provided in this particular hierarchy were: (a) the superordinate class of the target word, (b) definition, (c) a closure sentence, and (d) auditory target comprehension (see Abel, Willmes, & Huber, 2007, from which this semantic cue hierarchy was based on). Effects were analysed at the group level (n = 14) across therapies (participants also went through therapy using phonological cues) and collectively, there was significant improvement on the treated items when assessed one week after therapy (p. 159).
Provision of phonological cue(s): Explicit Scale of measurement: Binary – Yes, No	Whether phonological cue(s) is/are explicitly provided during the actual naming task. Examples include: Presenting the first sound or syllable in the spoken form. Compulsory repetition is not required. This variable excludes whole word presentation, which is coded as a separate variable. It also excludes phonological cues	Biedermann and Nickels (2008): Therapy relied on presenting cues in progressive order that convey phonological information exclusively. In this case, the cues were: (a) initial phoneme cue (consonant + schwa or vowel) and (b) tapping the number of syllables in the target word.

[4] Exactly how a hierarchy of semantic cues should look like is arguably a grey area, with unclear evidence that prioritises one semantic cue over another. In this case, the hierarchy presumably follows the amount of information provided, with the 'superordinate' cue being the vaguest or least meaningful.

	generated and/or provided during Phonological Component Analysis (PCA). PCA is a multi-step program, distinct from the basic direct application of phonological cues (Leonard, Rochon, & Laird, 2008).	Participant FME showed significant improvement on treated items (e.g., 'cricket' (game)) when assessed one day and one week after therapy. The researchers also distinguished between the types of untreated items. Significant improvement was observed for untreated homophone items (e.g., 'cricket (animal)) when assessed one day and one week after therapy. However, there was no improvement on untreated items that were not homophones (e.g., 'ant')[5] on both assessment points after therapy.
Provision of orthographic cue(s): Explicit Scale of measurement: Binary – Yes, No	Whether orthographic cue(s) is/are explicitly provided during the actual naming task. Examples include: Presenting the first letter or syllable in the written form. Compulsory copying or reading aloud is not required. This variable excludes whole word presentation, which is coded as a separate variable. It also excludes part-word presentation for copying.	Lorenz and Nickels (2007): In one of their experimental cue conditions, just the initial letter of the target word was presented to facilitate naming. For all three participants, the authors documented instances of significant improvements on treated items when cued with the initial letter, be it assessed immediately or a day after the facilitation session. In one participant, MCB, the effects were even maintained one week later (p. 680).

[5] The second participant in their study, KCC, showed a slightly different picture, with significantly improved naming on untreated homophones observed one day after treatment, but this was not maintained one week later. Conversely, KCC did not show improved naming on treated items one day after treatment, but showed marginally significant improvement one week later (p = .083; see pp. 285-286 in Biedermann and Nickels, 2008, for details). Also note that this is a rare instance in which efforts were made to distinguish between the different types of untreated words.

Provision of phonological cue(s): Implicit Scale of measurement: Binary – Yes, No	Whether phonological cue(s) is/are implicitly provided during the actual naming task. An example includes contextual priming by way of grouping phonologically-related target words together during naming.	Hendricks, Nicolas, and Zipse (2014): In this single-case study, therapy involved having the participant to name phonologically-related words in triplets. Participant AH showed strongest improvement in the "front-matched triplets, high phonological neighbourhood density" treatment condition (post-therapy assessments were done at four time-points: shortly after treatment to 10 weeks afterwards). In this condition, the trio of target words shares the same initial sounds (phonologically-related) and each target belongs to a dense phonological neighbourhood.
Provision of the target word in spoken form Scale of measurement: Binary – Yes, No	Whether the whole target word is provided aurally at any stage of the therapy. Provision of the target word does not necessary require spoken repetition.	Best, Greenwood, Grassly, and Hickin (2008): The therapy was divided into two parts: (1) A single-word naming therapy, and (2) use of single words in connected speech. The former is relevant here. It included the presentation of the entire word in its spoken form, when the participant ultimately fails to produce it despite progressive cues provided. It is not mandatory for participants to repeat the word. All participants (*n* = 8) showed significant improvement when assessed shortly after the single-word naming therapy (pp. 396-398).

Provision of the target word in written form Scale of measurement: Binary – Yes, No	Whether the whole target word is provided in written form at any stage of the therapy. Provision of the target word does not necessary require reading aloud or copying.	Silkes, Dierkes and Kendall (2013): In this single-case study, therapy involved masked orthographic repetition priming. Written target words were used as the primes, and paired with their corresponding pictures. It is not compulsory to read aloud the words, if seen. Mixed findings were found for the two lists used in treatment. For the "animal" list, no substantial effects were found for both treated and untreated items. For the "things to wear" list, a medium effect size for treated items and a small effect for untreated items were respectively found, when assessed immediately after the treatment ("A2 vs first 7 baseline points in A1" in Silke et al.'s Table 3; pp. 390-392).
Spoken repetition of target word: Compulsory Scale of measurement: Binary – Yes, No	Participant is obliged to repeat the given target word verbally, independent of whether the naming response (if required) is correct/incorrect during the therapy. Compulsory repetition can occur at any stage during therapy. The entire target word can be presented in the spoken and/or written form(s).	Mason, Nickels, McDonald, Moses, Makin and Taylor (2011): For therapy, the participants (*n* = 3) were shown a coloured-photograph along with its corresponding sound recording in each trial. The participant was instructed to repeat verbally after the recording. There were no other therapy technique components involved. Two participants (JMM, DRS) did not comply with the therapy instructions. For the participant who complied, SJS, significant improvement was observed for only one set

		of items (Set A) when assessed one week later, and these treatment-specific gains were maintained when assessed against the average baseline scores three weeks later, after the entire course of therapy ended. No generalization to untreated items was found.
Spoken repetition of target word: Contingent on response Scale of measurement: Binary – Yes, No	The participant is only required to repeat the given target word verbally, when the naming response is incorrect. The entire target word can be presented in the spoken and/or written form(s).	Altmann, Hazamy, Carvajal, Benjamin, Rosenbek, and Crosson (2014): In Phases 1 and 2 of the therapy, a picture was presented for naming. For correct responses, the participant will proceed to the next trial. For incorrect responses, the therapist will provide the spoken form of the target word for the participant to verbally repeat. Group analyses (*n* = 14) showed that collectively, compared to baseline performance, participants significantly improved on their naming of treated and untreated items when probed immediately post-therapy and three-months later (p. 446)[6].
Copying of written target word: Compulsory Scale of measurement: Binary – Yes, No	The participant is obliged to copy (orthographically) the given target word, independent of whether the naming response (if required) is correct/incorrect. This	Weill-Chounlamountry, Capelle, Tessier and Pradat-Diehl (2013): In this single-case intervention, part of the therapy involved the compulsory copying of

[6] There was an additional manipulation, with the 14 participants equally divided into the gesture and no-gesture groups. The only difference is that for the gesture group, a hand movement must be made to initiate the treatment trials (i.e., use the left hand to reach for box and press a red button) or when making naming corrections (left hand must make a non-meaningful circular gesture). This manipulation did not result in any interaction effects (see p. 444, in Altman et al., 2014).

	compulsory act of copying can occur at any stage during therapy. The entire target word is presented in the written form.	the written target word in presence of target picture. Participant HA significantly improved on the naming of trained and untrained words (*n* = 90) when assessed after therapy (time frame not specified, presumably shortly). This improvement was maintained two months later.
Copying of written target word: Contingent on response Scale of measurement: Binary – Yes, No	The participant is only required to copy (orthographically) the given target word, when the naming response is incorrect. The entire target word is presented in the written form.	Wright, Marshall, Wilson, and Page (2008): Therapy included having the participants (*n* = 2) copy down written target words, which (s)he failed to name and write. Participant P2 improved significantly on the naming of all sets of trained items when assessed shortly after treatment and one-month later. Significant improvement was also observed for untreated items when assessed one-month later. Participant P1 only showed significant improvement on the naming of all sets of trained items when assessed shortly after treatment.
Semantic cue(s) generation Scale of measurement: Binary – Yes, No	The participant is required to generate semantic cues. The cues can be generated in spoken and/or written form(s).	Marcotte, Adrover-Roig, Damien, de Préaumont, Généreux, Hubert, Ansaldo (2012): Therapy involved modified Semantic Feature Analysis. Participants (*n* = 9) were guided to generate semantic features related to the target, e.g., the group it belongs to, its use and location.

		On balance, the nine participants obtained a mean improvement of 80% on the trained items, and “some degree of generalization … to untrained materials” (p. 1780).
Phonological cue(s) generation Scale of measurement: Binary – Yes, No	Participant is required to generate phonological cues. The cues can be generated in spoken and/or written form(s).	van Hees, Angwin, McMahon, and Copland (2013): One of the therapies involved phonological component analysis (Leonard et al., 2008). Participants (*n* = 8) were guided to generate the first sound, syllable, last sound, association, and rhyme of the target word. Seven participants improved significantly on the naming of trained items when assessed immediately post-therapy and six maintained the gains two to three weeks later. No generalization was observed to untrained items.
Semantic task(s) Scale of measurement: Binary – Yes, No	The participant performs tasks that focus on the meaning of the target words. Examples include word-to-picture matching task with semantic distractors, semantic categorisation task, or semantic odd-one-out task. In these tasks, the semantic relatedness of the target to other items is typically explored. This variable excludes having the participant generate semantic cues (e.g., Semantic Feature Analysis), as that is coded separately.	Jacquemot, Dupoux, Robotham, and Bachoud-Lévi (2012): Therapy in Phase 1 of this single-case intervention focused on semantics and included semantic categorisation (labelled as “picture categorisation” here), function-matching (selecting the picture that best represents the function of the target item), semantic-verification (labelled as “knowledge assessment”) tasks.

		Participant DPI improved on his naming of the trained items after therapy in Phase 1, but this improvement was not significant.
Orthographic task(s) Scale of measurement: Binary – Yes, No	The participant performs tasks that focus on the spellings of the target words. Examples include anagrams, first letter or syllable identification. In these tasks, letter identification and letter sequencing of the target word are typically explored.	Weill-Chounlamountry, Capelle, Tessier and Pradat-Diehl (2013): Part of the therapy involved the rearranging jumbled letters in this single-case study. Participant HA significantly improved on the naming of trained and untrained words. This improvement was maintained two months later.

Table 5. Information or variables describing the application of technique(s).

Variable	Operationalization	Examples from Literature
Application of cues: Progression – Increasing vs. Decreasing Scale of measurement: Category – (1) Increasing, (2) Decreasing, (3) Not applicable	Whether the part-word cues are applied in an increasing (from small to bigger units; e.g., single phoneme to syllable) or decreasing (from big to smaller units; e.g., syllable to single phoneme) progression.	Conroy, Sage, and Lambon Ralph (2009d) compared the two approaches of cueing (increasing versus decreasing). At the individual level, all participants (*n* = 7) benefitted from both approaches, as each approach yielded significant improvements on naming for their respective set of treated items. The significant improvements were observed when assessed immediately at one-week post-therapy and at follow-up, five weeks after therapy ended.
Application of cues: Contingent on response vs. compulsory regardless of response (shortened as "Application of cues: Contingent on response" in Figures 2 to 3) Scale of measurement: Category – (1) Contingent, (2) Compulsory application, (3) Mixture of both, (4) Not applicable	Whether the cues are applied according to the response or independent of the response.	In Biedermann and Nickels (2008; also reported in Table 4), we see the application of cues that are contingent on the accuracy of the response. Cues were progressively applied from initial phoneme to whole word repetition, if responses were incorrect or not made within 10 seconds. Harnish et al. (2014) illustrates the obligatory application of cues irrespective of naming accuracy. There were eight steps (including different cues at different times) in their therapy, which "progressed through each of the eight presentations regardless of the participants' responses … This provided opportunity for multiple repetitions of the correct responses." (p. S288). The authors argued that all participants (*n* = 8) showed treatment-specific gains in naming post-

		therapy, and out of the seven assessed 8 weeks after follow-up, six maintained their gains (see p. S291 in their original paper). There was no reliable evidence of generalization to untreated items. Conroy and Scowcroft (2012) is an example of a mixture of cues. The cueing process begins as compulsory whole word repetition and then it becomes contingent on the participant's response – i.e., if the participant fails to name when picture was presented for a second time, the whole word is presented again; else the experimenter proceeds to the next round using CV syllable as cues (see p. 307 of their paper for a description). All participants (*n* = 4) in Conroy and Scowcroft (2012) showed significant improvements on their naming of treated items post-therapy, and these were also significantly higher than scores posted for the untreated control items.
Feedback on accuracy of the naming response Scale of measurement: Category – (1) No feedback, (2) explicit feedback provided, (3) implicit feedback provided	The type of feedback provided to the participant about naming accuracy. This is for each of the naming response made during therapy. There can be (1) no feedback, (2) explicit feedback, or (3) implicit feedback. Explicit feedback refers to instances in which at all stages of therapy where naming responses are needed, there is direct	Off, Griffin, Spencer, and Rogers (2016) is an example of a therapy with nil feedback. "Participants did not receive feedback at any time during training. That is, participants did not have information about whether they accurately produced the correct name for the target." (p. 993) Macoir, Routhier, Simard and Picard (2012) is an example of a therapy with explicit feedback.

	acknowledgement of accurate responses and pointing out of inaccurate responses. Implicit feedback refers to instances in which at any stage in the therapy where naming responses are needed, direct acknowledgement of the accuracy of response is not present, but can be inferred. Typically, this takes the form of progressing to next naming trial only if the response is correct. Cues, however, are provided if the response is incorrect.	"Feedback was given for each response … When an error was produced, the patient was given the correct answer, along with an explanation about the nature of the error." (p. 124) Best et al. (2008; also reported in Table 4) is an instance of a therapy with implicit feedback. Although feedback was not directly provided to the participants on the accuracy of their responses, but they could infer based on whether the experimenter/therapist proceeded to the next trial or supplied them with a cue.
Termination criteria Scale of measurement: Category – (1) Performance criterion; (2) Fixed number of sessions completed; (3) Mix of performance criterion and certain number of sessions completed	Therapy can be broadly terminated in three ways: (1) The participant manages to achieve a certain level of performance (i.e., criteria present); (2) The participant has completed a fixed number of sessions (i.e., criteria absent); or (3) The participant meets a certain level of performance or manages to complete a fixed number of sessions (i.e., mix of performance criterion and a fixed number of sessions)	Rider, Wright, Marshall, and Page (2008) is an example of a therapy with performance criteria. The participant moved on to the next stage of the therapy (i.e., names next list of items in this case), after achieving 80% naming accuracy, or having completed 10 trials (p. 164). Conroy, Sage, and Lambon Ralph (2009a) is an example of a therapy without performance criteria. The participants were simply required to complete 10 sessions of therapy (p. 474). In Marcotte, Adrover-Roig, Damien, de Preaumont, Genereux, Hubert, and Ansaldo (2012), therapy was terminated when participants achieved 80% performance on

		treated items or completed a maximum of six weeks of therapy (p. 1778).
Mode in which the actual therapy was delivered Scale of measurement: Category - (1) Machine, (2) Individual, (3) Combination of machine and individual	Whether the actual therapy is delivered by: (1) the machine (i.e., computer-based/assisted or self-administered therapy such that the computer is providing the cues, for instance; the use of computer can be unsupervised, supervised or assisted), (2) a person (the therapist / experimenter/ trained individual is providing the cues) or (3) a combination of machine and person (e.g., orthographic cues presented on the computer, but phonological cues are supplied by the therapist; or some parts of therapy are completely computerised but other parts are solely provided by the therapist)	Adrian, Gonzalez, Buiza, and Sage (2011) pilot-tested the Spanish Computer-assisted Anomia Rehabilitation Program (CARP). It is an example of a computer providing the actual therapy cues. All their participants (*n* = 15) improved significantly on the naming of treated items and 11 of them showed generalization to untreated item post-treatment (seven to ten days after the treatment stage)(also see Routhier, Bier, & Macoir, 2016, for a self-administered word-finding therapy on a tablet). Harnish et al. (2014; see above) is an example of a therapy that uses a mixture of both computer and therapist to provide therapy. The orthographic cues were presented on the computer screen, but the phonological and semantic cues were spoken aloud by the therapist (see p. S285 in the original paper).
Homework Scale of measurement: Binary – Yes, No	Whether homework is assigned as part of the therapy package	In Conroy and Scowcroft (2012), we find an example in which homework was assigned. Participants (*n* = 4; likely with assistance of caregivers) were asked to look at the pictures and read aloud the corresponding written words underneath three times. Unfortunately, despite explicit instructions requesting participants to document their frequencies of homework completion, no

participants documented so with sufficient consistency to permit analysis on the relationship between homework and naming progress (see p. 315 of the original paper).

Table 6. Descriptive statistics based on imputed values generated by the representative random forest for naming performance on treated items, when assessed at the short-term phase after therapy.

Name of variable	Mean (*SD*) or frequency count of cases, where applicable	
	Improved outcome observed (*n* = 167)	**Nil improvement in outcome observed (*n* = 17)**
1. Participant-related variables		
Age (Years)	61.35 (*SD* = 11.62)	56.12 (*SD* = 11.71)
Gender	Male: 88; Female: 79	Male: 10; Female: 7
Months post-onset of stroke	41.29 (*SD* = 45.26)	76.53 (*SD* = 102.24)
Years of education received	11.87 (*SD* = 3.48)	11.93 (*SD* = 1.98)
Boston Naming Test score at baseline (Percent)	0.41 (*SD* = 0.21)	0.49 (*SD* = 0.25)
2. Variables describing the regimen		
Number of minutes per session	56.63 (*SD* = 7.98)	63.84 (*SD* = 12.45)
Number of sessions per week	2.17 (*SD* = 0.61)	2.97 (*SD* = 0.59)
Number of sessions for the entire therapy	13.94 (*SD* = 7.80)	12.71 (*SD* = 3.30)
3. Variables describing the items (target words)		
Total number of items treated across sessions	46.12 (*SD* = 42.93)	38.24 (*SD* = 13.85)
Total number of times each item was named across sessions	70. 66 (*SD* = 27.53)	31.87 (*SD* = 15.79)
Grammatical class of the treated words	Nouns: 42; Verbs: 64; Mixture of nouns and verbs: 61	Nouns: 5; Verbs: 10; Mixture of nouns and verbs: 2
4. Variables describing the techniques		
Provision of semantic cue(s): Explicit	Yes: 97; No: 70	Yes: 5; No: 12
Provision of phonological cue(s): Explicit	Yes: 145; No: 22	Yes: 6; No: 11
Provision of orthographic cue(s): Explicit	Yes: 149; No: 18	Yes: 2; No: 15
Provision of phonological cue(s): Implicit	Yes: 1; No: 166	No: 17
Provision of the target word in spoken form	Yes: 167	Yes: 13; No: 4
Provision of the target word in written form	Yes: 151; No: 16	Yes: 2; No: 15

Spoken repetition of target word: Compulsory	Yes: 55; No: 112	Yes: 1; No: 16
Spoken repetition of target word: Contingent on response	Yes: 147; No: 20	Yes: 13; No: 4
Copying of written target word: Compulsory	Yes: 1; No: 166	No: 17
Copying of written target word: Contingent on response	No: 167	No: 17
Semantic cue(s) generation	Yes: 3; No: 164	Yes: 5; No: 12
Phonological cue(s) generation	Yes: 6; No: 161	Yes: 2; No: 15
Semantic task(s)	Yes: 32; No: 135	No: 17
Orthographic task(s)	Yes: 1; No: 166	No: 17
5. Variables describing the application of technique(s)		
Application of cues: Progression	Increasing: 102; Decreasing: 36; Not applicable: 29	Increasing: 5; Not applicable: 12
Application of cues: Contingent on response	Contingent on response: 85; Compulsory: 19; Mixture of compulsory and response-contingent applications: 52; Not applicable: 11	Contingent on response: 5; Compulsory: 1; Not applicable: 11
Feedback on accuracy of the naming response	Explicit feedback: 6; Implicit feedback: 153; No feedback: 8	Explicit feedback: 4; Implicit feedback: 6; No feedback: 7
Termination criteria	Performance criterion: 1; Fixed number of sessions: 166	Fixed number of sessions: 17
Mode in which actual therapy was delivered	Computer: 62; Person: 105	Computer: 2; Person: 15
Homework	Yes: 8; No: 159	No: 17

Table 7. Descriptive statistics based on imputed values generated by the representative random forest for naming performance on treated items, when assessed during the maintenance phase after therapy.

Name of variable	Mean (*SD*) or frequency count of cases, where applicable	
	Improved outcome observed (*n* = 116)	**Nil improvement in outcome observed (*n* = 11)**
1. Participant-related variables		
Age (Years)	62.63 (*SD* = 11.44)	59.55 (*SD* = 14.36)
Gender	Male: 69; Female: 47	Male: 10; Female: 1
Months post-onset of stroke	48.90 (*SD* = 60.43)	133.45 (*SD* = 177.24)
Years of education received	13.43 (*SD* = 1.58)	13.49 (*SD* = 1.86)
Boston Naming Test score at baseline (Percent)	0.40 (*SD* = 0.22)	0.33 (*SD* = 0.11)
2. Variables describing the regimen		
Number of minutes per session	53.65 (*SD* = 8.40)	61.82 (*SD* = 14.71)
Number of sessions per week	2.18 (*SD* = 0.66)	3.15 (*SD* = 0.62)
Number of sessions for the entire therapy	10.30 (*SD* = 2.29)	12.09 (*SD* = 4.29)
3. Variables describing the items (target words)		
Total number of items treated across sessions	30.85 (*SD* = 15.75)	22.91 (*SD* = 14.52)
Total number of times each item was named across sessions	69.03 (*SD* = 29.12)	51.50 (*SD* = 21.32)
Grammatical class of the treated words	Nouns: 26; Verbs: 40; Mixture of nouns and verbs: 50	Nouns: 6; Verbs: 4; Mixture of nouns and verbs: 1
4. Variables describing the techniques		
Provision of semantic cue(s): Explicit	Yes: 49; No: 67	Yes: 1; No: 10
Provision of phonological cue(s): Explicit	Yes: 97; No: 19	Yes: 2; No: 9
Provision of orthographic cue(s): Explicit	Yes: 107; No: 9	Yes: 3; No: 8
Provision of phonological cue(s): Implicit	Yes: 1; No: 115	No: 11
Provision of the target word in spoken form	Yes: 115; No: 1	Yes: 7; No: 4
Provision of the target word in written form	Yes: 111; No: 5	Yes: 3; No: 8

Spoken repetition of target word: Compulsory	Yes: 50; No: 66	Yes: 2; No: 9
Spoken repetition of target word: Contingent on response	Yes: 91; No: 25	Yes: 5; No: 6
Copying of written target word: Compulsory	Yes: 1; No: 115	No: 11
Copying of written target word: Contingent on response	Yes: 1; No: 115	No: 11
Semantic cue(s) generation	Yes: 3; No: 113	Yes: 1; No: 10
Phonological cue(s) generation	No: 116	No: 11
Semantic task(s)	No: 116	No: 11
Orthographic task(s)	Yes: 2; No: 114	No: 11
5. Variables describing the application of technique(s)		
Application of cues: Progression	Increasing: 63; Decreasing: 29; Not applicable: 24	Increasing: 3; Not applicable: 8
Application of cues: Contingent on response	Contingent on response: 83; Compulsory: 18; Mixture of compulsory and response-contingent applications: 8; Not applicable: 7	Contingent on response: 3; Compulsory: 1; Not applicable: 7
Feedback on accuracy of the naming response	Explicit feedback: 1; Implicit feedback: 109; No feedback: 6	Explicit feedback: 3; Implicit feedback: 2; No feedback: 6
Termination criteria	Performance criterion: 1; Fixed number of sessions: 109; Mixture of the two: 6	Fixed number of sessions: 9 Mixture of the two: 2
Mode in which actual therapy was delivered	Computer: 6; Person: 110	Computer: 3; Person: 8
Homework	Yes: 8; No: 108	No: 11

Table 8. Descriptive statistics based on imputed values generated by the representative random forest for naming performance on untreated items, when assessed at the short-term phase after therapy.

Name of variable	Mean (*SD*) or frequency count of cases, where applicable	
	Improved outcome observed (*n* = 27)	Nil improvement in outcome observed (*n* = 125)
1. Participant-related variables		
Age (Years)	59.67 (*SD* = 13.88)	62.25 (*SD* = 11.12)
Gender	Male: 9; Female: 18	Male: 74; Female: 51
Months post-onset of stroke	31.63 (*SD* = 25.05)	48.20 (*SD* = 61.14)
Years of education received	9.14 (*SD* = 5.40)	12.55 (*SD* = 2.72)
Boston Naming Test score at baseline (Percent)	0.40 (*SD* = 0.11)	0.43 (*SD* = 0.23)
2. Variables describing the regimen		
Number of minutes per session	58.29 (*SD* = 4.67)	59.81 (*SD* = 8.35)
Number of sessions per week	2.30 (*SD* = 0.22)	2.35 (*SD* = 0.55)
Number of sessions for the entire therapy	26.30 (*SD* = 7.92)	11.80 (*SD* = 5.16)
3. Variables describing the items (target words)		
Total number of items treated across sessions	85.19 (*SD* = 75.91)	35.87 (*SD* = 28.40)
Total number of times each item was named across sessions	72.63 (*SD* = 19.33)	71.06 (*SD* = 29.84)
Grammatical class of the treated words	Nouns: 11; Verbs: 13; Mixture of nouns and verbs: 3	Nouns: 15; Verbs: 43; Mixture of nouns and verbs: 67
4. Variables describing the techniques		
Provision of semantic cue(s): Explicit	Yes: 23; No: 4	Yes: 64; No: 61
Provision of phonological cue(s): Explicit	Yes: 27	Yes: 94; No: 31
Provision of orthographic cue(s): Explicit	Yes: 27	Yes: 100; No: 25
Provision of phonological cue(s): Implicit	No: 27	No: 125
Provision of the target word in spoken form	Yes: 27	Yes: 121; No: 4
Provision of the target word in written form	Yes: 27	Yes: 80; No: 45

Spoken repetition of target word: Compulsory	Yes: 2; No: 25	Yes: 37; No: 88
Spoken repetition of target word: Contingent on response	Yes: 27	Yes: 109; No: 16
Copying of written target word: Compulsory	No: 27	Yes: ; No: 125
Copying of written target word: Contingent on response	No: 27	No: 125
Semantic cue(s) generation	No: 27	Yes: 8; No: 117
Phonological cue(s) generation	No: 27	Yes: 8; No: 117
Semantic task(s)	Yes: 22; No: 5	Yes: 10; No: 115
Orthographic task(s)	No: 27	No: 125
5. Variables describing the application of technique(s)		
Application of cues: Progression	Increasing: 25; Decreasing: 2	Increasing: 68; Decreasing: 26; Not applicable: 31
Application of cues: Contingent on response	Contingent on response: 5; Mixture of compulsory and response-contingent applications: 22	Contingent on response: 70; Compulsory: 11; Mixture of compulsory and response-contingent applications: 22; Not applicable: 22
Feedback on accuracy of the naming response	Implicit feedback: 27	Explicit feedback: 2; Implicit feedback: 100; No feedback: 23
Termination criteria	Fixed number of sessions): 27	Performance criterion: 1; Fixed number of sessions: 124
Mode in which actual therapy was delivered	Computer: 22; Person: 5	Computer: 12; Person: 113
Homework	No: 27	No: 125

Figure 1. PRISMA flow chart for systematic review (Moher, Liberati, Tetzlaff, Altman, and The PRISMA Group, 2009).

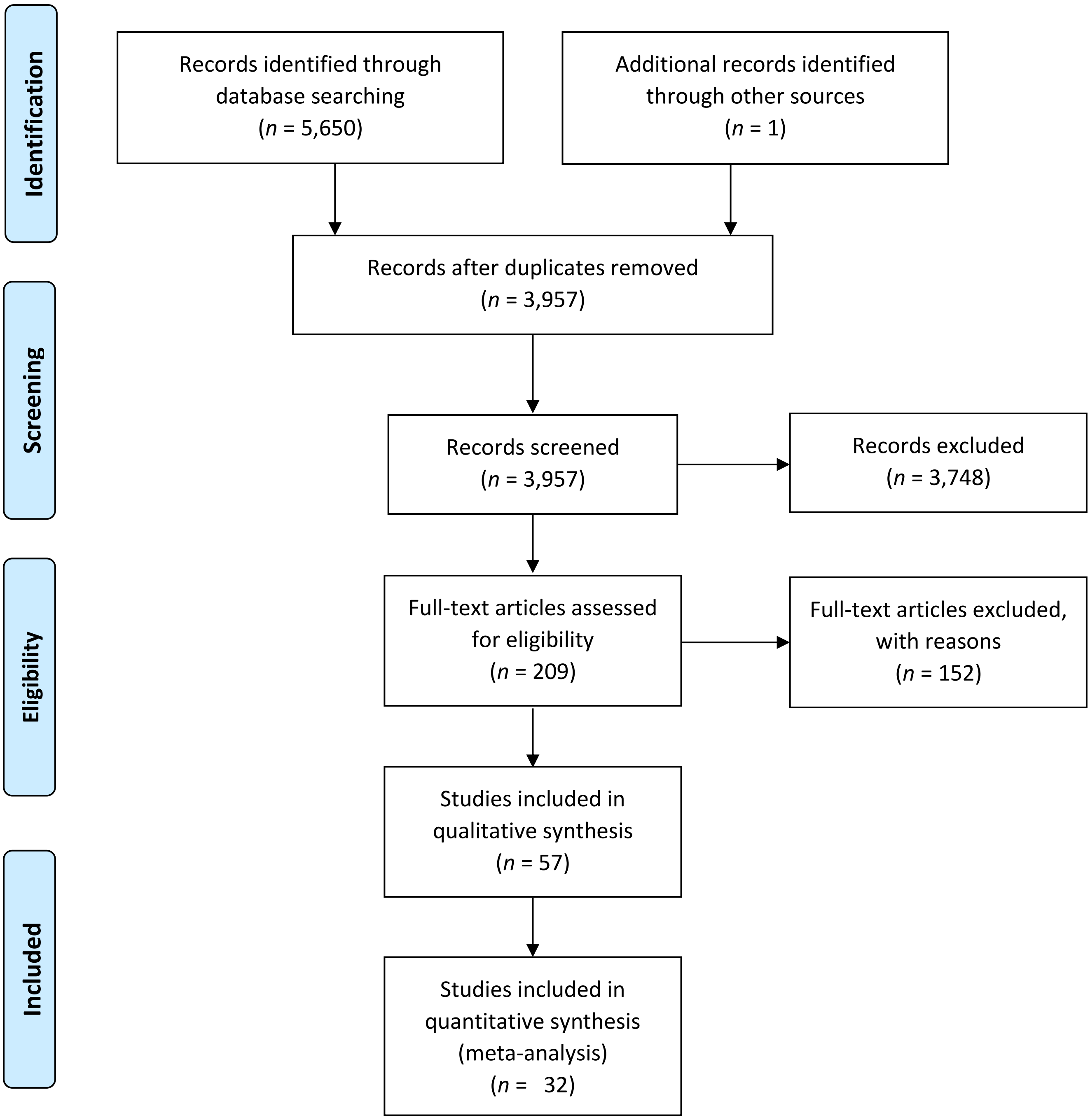

Identification
Screening
Eligibility
Included
Records identified through database searching (n = 5,650)
Additional records identified through other sources (n = 1)
Records after duplicates removed (n = 3,957)
Records screened (n = 3,957)
Records excluded (n = 3,748)
Full-text articles assessed for eligibility (n = 209)
Full-text articles excluded, with reasons (n = 152)
Studies included in qualitative synthesis (n = 57)
Studies included in quantitative synthesis (meta-analysis) (n = 32)

Figure 2. Variables and their values of variable importance for improved lexical-retrieval of treated items post-therapy in the short-term phase.

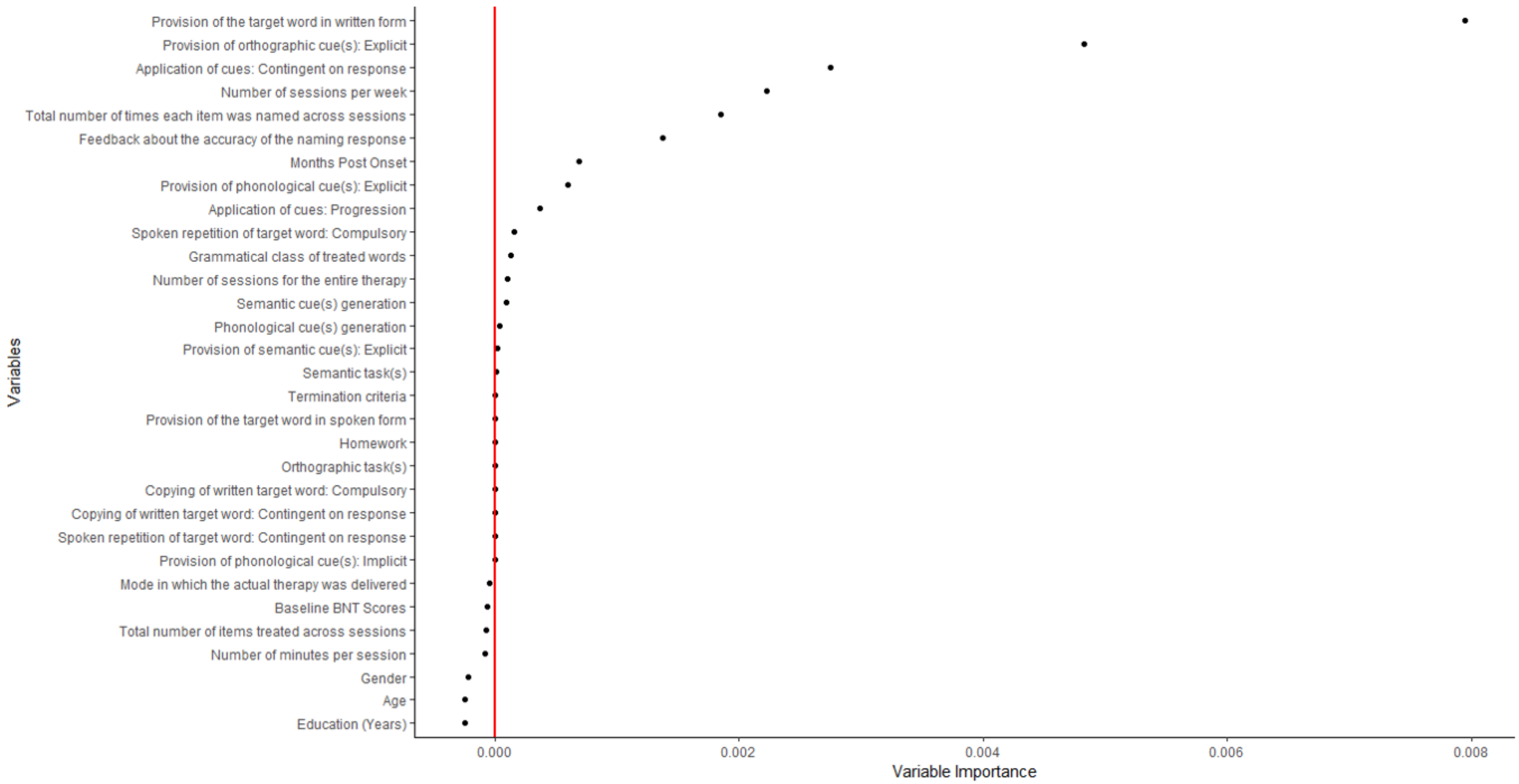
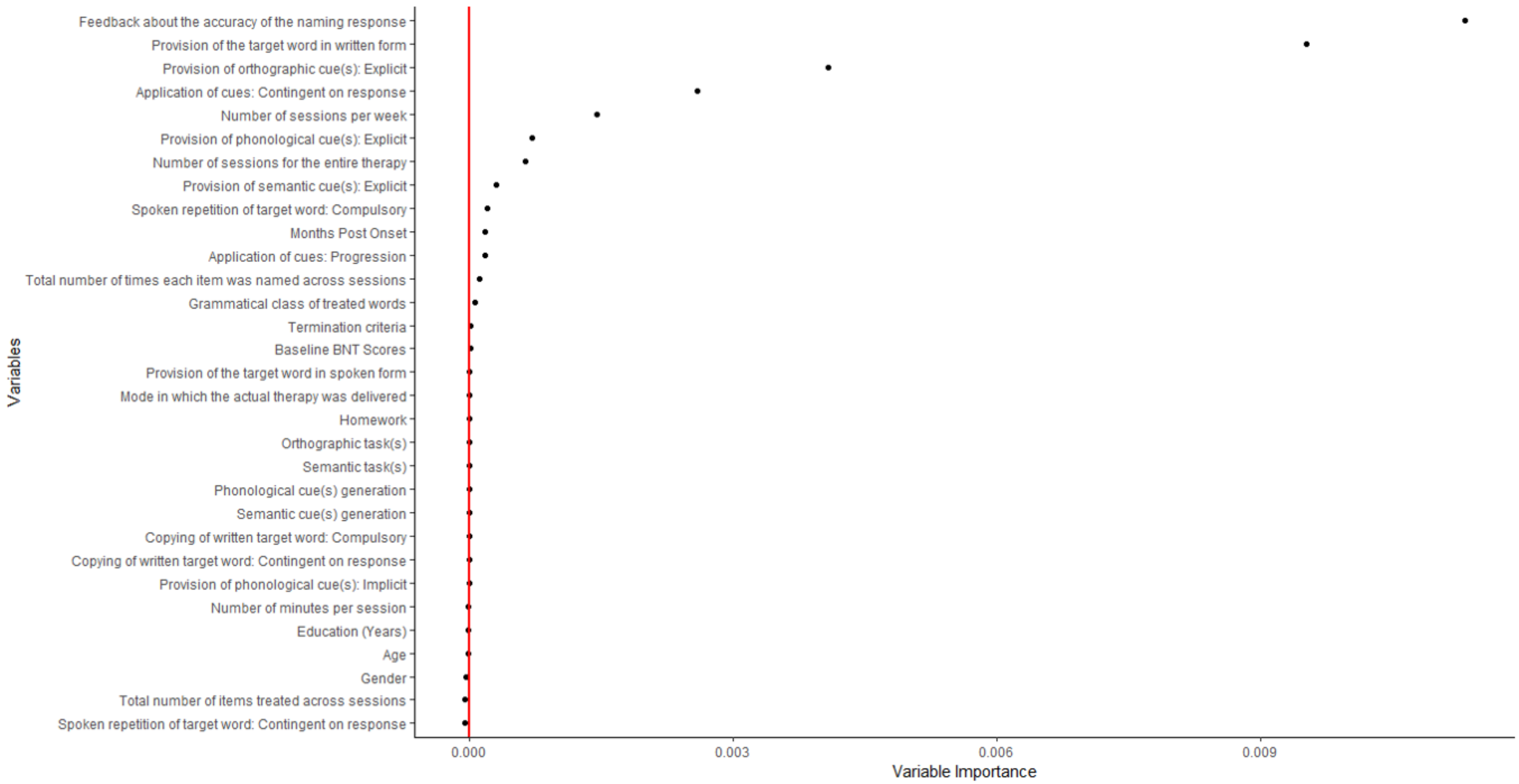
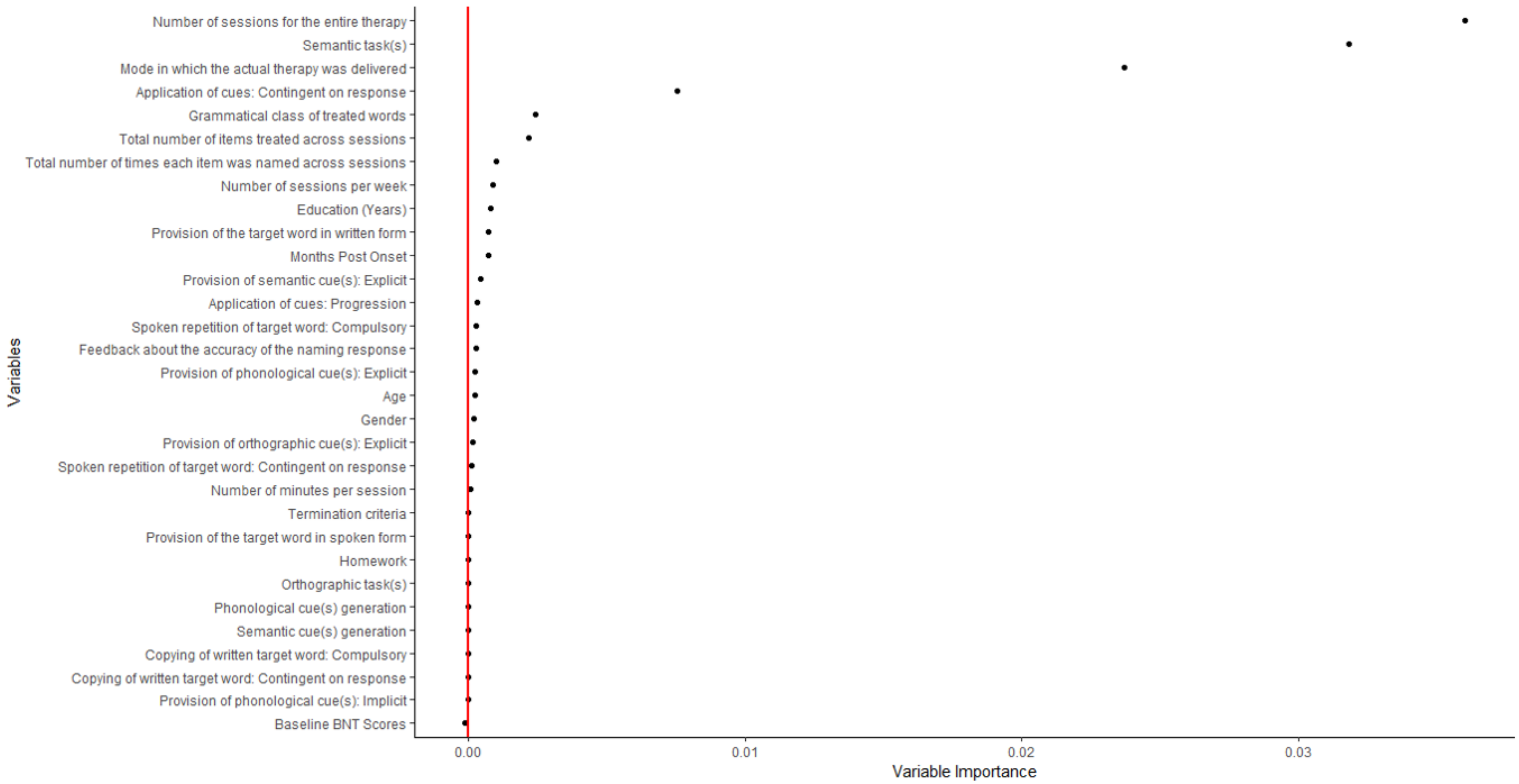

Provision of the target word in written form
Provision of orthographic cue(s): Explicit
Application of cues: Contingent on response
Number of sessions per week
Total number of times each item was named across sessions
Feedback about the accuracy of the naming response
Months Post Onset
Provision of phonological cue(s): Explicit
Application of cues: Progression
Spoken repetition of target word: Compulsory
Grammatical class of treated words
Number of sessions for the entire therapy
Semantic cue(s) generation
Phonological cue(s) generation
Provision of semantic cue(s): Explicit
Semantic task(s)
Termination criteria
Provision of the target word in spoken form
Homework
Orthographic task(s)
Copying of written target word: Compulsory
Copying of written target word: Contingent on response
Spoken repetition of target word: Contingent on response
Provision of phonological cue(s): Implicit
Mode in which the actual therapy was delivered
Baseline BNT Scores
Total number of items treated across sessions
Number of minutes per session
Gender
Age
Education (Years)
Variables
0.000
0.002
0.004
0.006
0.008
Variable Importance

Figure 3. Variables and their values of variable importance for improved lexical-retrieval of treated items post-therapy during the maintenance phase.



Feedback about the accuracy of the naming response
Provision of the target word in written form
Provision of orthographic cue(s): Explicit
Application of cues: Contingent on response
Number of sessions per week
Provision of phonological cue(s): Explicit
Number of sessions for the entire therapy
Provision of semantic cue(s): Explicit
Spoken repetition of target word: Compulsory
Months Post Onset
Application of cues: Progression
Total number of times each item was named across sessions
Grammatical class of treated words
Termination criteria
Baseline BNT Scores
Provision of the target word in spoken form
Mode in which the actual therapy was delivered
Homework
Orthographic task(s)
Semantic task(s)
Phonological cue(s) generation
Semantic cue(s) generation
Copying of written target word: Compulsory
Copying of written target word: Contingent on response
Provision of phonological cue(s): Implicit
Number of minutes per session
Education (Years)
Age
Gender
Total number of items treated across sessions
Spoken repetition of target word: Contingent on response
Variables
0.000
0.003
0.006
0.009
Variable Importance

Figure 4. Variables and their values of variable importance for improved lexical-retrieval of untreated items post-therapy in the short-term phase.


Number of sessions for the entire therapy
Semantic task(s)
Mode in which the actual therapy was delivered
Application of cues: Contingent on response
Grammatical class of treated words
Total number of items treated across sessions
Total number of times each item was named across sessions
Number of sessions per week
Education (Years)
Provision of the target word in written form
Months Post Onset
Provision of semantic cue(s): Explicit
Application of cues: Progression
Spoken repetition of target word: Compulsory
Feedback about the accuracy of the naming response
Provision of phonological cue(s): Explicit
Age
Gender
Provision of orthographic cue(s): Explicit
Spoken repetition of target word: Contingent on response
Number of minutes per session
Termination criteria
Provision of the target word in spoken form
Homework
Orthographic task(s)
Phonological cue(s) generation
Semantic cue(s) generation
Copying of written target word: Compulsory
Copying of written target word: Contingent on response
Provision of phonological cue(s): Implicit
Baseline BNT Scores
Variables
0.00
0.01
0.02
0.03
Variable Importance