A Systematic Review of Semantic Feature Analysis Therapy Studies for Aphasia

E. A. Efstratiadou [1,2], I. Papathanasiou [2,3], R. Holland[1], A. Archonti[2], K. Hilari[1]

[1]Division of Language and Communication Science, City, University of London, UK

[2]Thales Aphasia Project, Department of Linguistics, School of Philosophy, University of Athens, Greece

[3]Department of Speech and Language Therapy, TEI of Western Greece Patras, Greece

Corresponding author

Professor Katerina Hilari

Division of Language and Communication Science

School of Health Sciences

City, University of London

Northampton Square,

London

EC1V 0HB

UK

k.hilari@city.ac.uk

Tel: +44 (0) 207 040 4660

This is the author version of this paper, which has been accepted for publication in the *Journal of Speech, Language and Hearing Research*

Abstract

Purpose: The purpose of this study was to review treatment studies of semantic feature analysis (SFA) for persons with aphasia. The review documents how SFA is used, appraises the quality of the included studies and evaluates the efficacy of SFA.

Methods: The following electronic databases were systematically searched (last search February 2017): Academic Search Complete; CINAHL Plus; E-journals; Health Policy Reference Centre; MEDLINE; PsycARTICLES; PsycINFO; and SocINDEX. The quality of the included studies was rated. Clinical efficacy was determined by calculating effect sizes (*Cohen's d*) or percent of non-overlapping data when *d* could not be calculated.

Results: Twenty-one studies were reviewed reporting on 55 persons with aphasia. SFA was used in six different types of studies: confrontation naming of nouns, of verbs, connected speech/discourse, group, multilingual and studies where SFA was compared with other approaches. The quality of included studies was high [Single Case Experimental Design Scale (SCEDS) average (range) =9.55 (8.0-11)]. Naming of trained items improved for 45 participants (81.82%). Effect sizes indicated there was a small treatment effect.

Conclusions: SFA leads to positive outcomes despite the variability of treatment procedures, dosage, duration and variations to the traditional SFA protocol. Further research is warranted to examine the efficacy of SFA and generalization effects in larger controlled studies.

Key words: Semantic feature analysis, Aphasia, Anomia, Treatment, Systematic review, efficacy

Introduction

A persistent and frequent symptom of aphasia is anomia, which is a difficulty or inability to find the right word (Goodglass & Wingfield, 1997). Anomia has been described as "the most consistent feature of aphasia" as virtually all people with aphasia experience some degree of word finding problems (Davis, 2000, p. 6). Being unable to find the right words impairs a person's ability to express their wants, needs, ideas and feelings, and participate in everyday conversations and social interactions. Reduced communicative participation in turn affects the person's emotional and social well-being and quality of life (Fotiadou, Northcott, Chatzidaki, & Hilari, 2014; Hilari, Needle, & Harrison, 2012; Northcott, Moss, Harrison, & Hilari, 2015).

Naming deficits in aphasia are very common. Naming requires processing at the level of word meaning (semantics), which connects to the word form (phonology) (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Goldrick, 2006; Levelt, 1999; Levelt, Roelofs, & Meyer, 1999;). Impairment in one or both of these processing stages, or the connections between them, can lead to difficulty in naming (Dell, Lawler, Harris, & Gordon, 2004; Levelt et al., 1999; Schwartz & Brecher, 2000; Schwartz, Dell, Martin, Gabl, & Sobel, 2006). Therapy for impaired naming can target semantic or phonological processing or a combination of these. Therapy approaches have used semantic, phonological and orthographic cues (Nickels, 2002; Wisenburn & Mahoney, 2009).

Semantic approaches aim to improve naming by restoring or strengthening semantic representations, or by priming weak semantic representations (Maher & Raymer, 2004). Semantic tasks described in the literature for improving naming in people with aphasia include: spoken and written word–picture matching (Byng, 1988; Marshall et al., 1990); generating semantic features of the object to be named - semantic feature analysis (Boyle, 2004; Boyle & Coelho, 1995; Coelho, McHugh, & Boyle, 2000; Lowell, Beeson, & Holland, 1995); semantic feature verification (Kiran & Thompson, 2003); generating or matching synonyms (Hough, 1993); contextual priming (Martin, Fink, & Laine, 2004; Renvall, Laine, & Martin, 2007); and making judgments about functions, semantic features, or relatedness of objects (Drew & Thompson, 1999; Nickels & Best, 1996a, 1996b).

Phonological approaches aim to strengthen representations at the level of the word form (Maher & Raymer, 2004), or strengthen the connections from the semantic system to the word form (Laine & Martin, 2006). Naming impairment due to deficits in post-semantic/phonological processing may be the result of impaired access to the phonological output lexicon, or in the lexical representations themselves (Laine & Martin, 2006). Phonological tasks include those that provide information about the phonology of the target (repetition, phonemic cues). Therapy tasks that have been shown to improve naming in people with aphasia include the use of cueing hierarchies and repetition (Raymer, Thompson, Jacobs, & Le Grand, 1993); reading aloud (Eales & Pring, 1998; Howard, 1994; Nickels & Best, 1996b), syllable judgments, initial phoneme discrimination, and rhyme judgment (Franklin, Buerk, & Howard, 2002; Robson, Marshall, Pring, & Chiat, 1998). Repetition is the most common phonological task, used in the majority of treatments (Nickels & Best, 1996; Nickels, 2002). A subset of phonological approaches has used orthographic cues, such as providing the first letter of the target word (Greenwood, Grassly, Hickin, & Best, 2010; Hickin, Best, Herbert, Howard, & Osborne, 2002; Leonard et al., 2004; Lorenz & Nickels, 2007). Lorenz & Nickels (2007) provide a detailed account of the mechanisms underlying the effectiveness of orthographic cues.

Traditionally, semantic and phonological tasks were thought to have different effects on word retrieval (Mitchum, Haendiges, & Berndt, 1995; Nickels & Best, 1996a, 1996b). Early research reported phonological tasks improved naming for a very short time, up to 10-15 minutes, whereas semantic tasks improved naming for up to 24 hours (Howard et al, 1985). However, more recent studies have shown that phonological cues can produce durable effects (Best & Nickels, 2000). Howard (2000) suggests that the difference between semantic and phonological tasks may well be overstated. As Howard (1994) and Nickels (2002a) indicated, most treatments comprise tasks that involve semantic, phonological, and sometimes orthographic tasks, despite the fact that researchers and clinicians typically characterize their treatments as either semantic or phonological. In the majority of the studies using semantic tasks, the form of the word is provided, as a spoken or written word, and/or repetition is required (suggesting phonological processing), and in phonological tasks, the picture is usually present (suggesting semantic processing). This is

also the case in Semantic Features Analysis (SFA) (Boyle & Coelho, 1995; Coelho et al., 2000; Conley & Coelho, 2003).

Ylvisaker and Szekeres (1985) were the first to introduce SFA as an organized method for facilitating semantic network activation. Ylvisaker, Szekeres and their colleagues provided general descriptions of SFA treatment, which emphasized the importance of using the structured procedure consistently. The approach was further developed and tested by Massaro and Tompkins (1994). They published the first data from a multiple-baseline, single-subject study of two individuals who had sustained traumatic brain injury. Theoretically, SFA is based on the concept of spreading activation within the semantic system (Collins & Loftus, 1975). Specifically, it was proposed that the level of semantic processing is conceptualized as a network of semantic representations and links to other related representations. Semantic representations with many shared properties were thought to link more closely together than representations that had minimal or no shared properties. The presentation of features that are strongly related to a target results in a spreading of activation that converges onto the target concept, which thus receives a higher level of activation than other similar concepts (Boyle, 2004; Boyle & Coelho, 1995; Coelho et al., 2000; Conley & Coelho, 2003; Haarbauer-Krupa, Moser, Smith, Sullivan, & Szekeres, 1985a & 1985b; Lowell, Beeson, & Holland, 1995; Massaro & Tompkins, 1994). For example, Boyle (2010) uses the example of “apple”. Its semantic features include <fruit >, <has a core>, <has skin>, <has seeds>, <grows on trees>, and <used for cider>. The information provided by its features differs, with some features providing more distinctive information (distinctive features) than others (common features). The feature <used for cider> distinguishes it from other fruits, like orange, whereas <has skin> does not distinguish it because most fruits have skin. The target concept then activates the phonological information associated with it, resulting in the production of the target word. SFA, therefore, relies upon re–learning, or applying a strategy you have learnt, which encourages activation between strongly associated features that in turn drives naming of a target picture or semantic concept (Hashimoto & Frome, 2011).

The SFA treatment protocol involves employing a “feature analysis chart” that includes the following semantic features for object naming: group, action, use, location, properties, and associations (Boyle, 2010) and for action naming: subject, purpose of action, part of body or tool used to carry out the action, description, usual location and associated objects or

actions (Wambaugh & Ferguson, 2007) (see Figure 1). During SFA treatment, individuals with word retrieval difficulties are shown a picture to name and they are encouraged to generate the semantic features of the target word by completing the feature analysis chart. The completion of the feature analysis chart is achieved by systematic cueing techniques, like asking questions or using sentence completion. For example, for 'rabbit', 'It is an…'('animal'), whilst pointing to the picture, 'It has ….' ('long ears / fluffy tail'), 'What does it do?' ('It hops'). The clinician guides the person with aphasia to complete the chart and gradually cueing is faded so that the person with aphasia becomes increasingly independent in generating features. It is argued that generation of such semantic features works as a compensatory strategy to enhance activation of the target word via the processing of shared features, which enables the individual to find the target word. Persistent and systematic practice in producing semantic features in this way enables individuals to achieve more organized word retrieval without the deliberate use of compensatory strategies (Boyle, 2010).

[figure 1 about here]

Two reviews have been previously conducted on SFA treatment. Boyle's (2010) report was the first and examined the efficacy of SFA. The review comprised seven studies where SFA was used for confrontation naming of nouns. Results were reported for 17 participants with aphasia, 16 of whom improved their ability to name pictured nouns. These participants had a variety of classic fluent and non-fluent aphasia syndromes. The review concluded that SFA treatments improve naming of treated items for most participants, regardless of whether they require participants to generate the features themselves or whether participants analyze features that have been generated by others (Boyle, 2010). Maddy, Capilouto and McComas (2014) conducted a systematic review on the same area, but excluded studies that involved verification rather than generation of features (Edmonds & Kiran, 2006; Kiran & Roberts, 2010). The review comprised 11 studies with 24 participants with aphasia. Seventeen of them had non-fluent aphasia and seven participants had fluent aphasia. *Cohen's d* was calculated and the majority of participants showed a small effect size. The percent of non-overlapping data was also calculated and a large treatment effect was present for the majority of participants. The review concluded that SFA is an effective intervention for improving confrontational naming of items trained in therapy; however, limited generalization to untrained items and connected speech was

reported in the majority of the included studies. The present study extends the previous reviews (Boyle, 2010; Maddy et al., 2014) in a number of ways. It includes new research. It evaluates the methodological quality of the existing studies against standard criteria (The Single Case Experimental Design Scale (SCEDS) critical appraisal tool (Tate et al., 2008) and level of evidence, based on the Scottish Intercollegiate Guidelines Network (http:sign.ac.uk/pdf/sign118.pdf, 2010). It also broadens the scope of the previous reviews by documenting the characteristics of SFA studies (participant characteristics, type of SFA, treatment dosage, treatment duration, total amount of treatment); and determining clinical efficacy. In particular, the following research questions were addressed:

1) What is the methodological quality of studies evaluating the efficacy of SFA in aphasia therapy? This will be rated against standard criteria.
2) What are the characteristics of SFA aphasia therapy studies, in terms of i) type, dosage, duration and total amount of treatment, and ii) participant characteristics?
3) What are the results of SFA aphasia therapy studies, in terms of i) treatment outcomes, and ii) clinical efficacy as determined by effect sizes using Cohen's d or percent of non-overlapping data?

Methods

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009 & 2010) formed the basis of the conduct and reporting of this systematic review. PRISMA stems from an international collaboration formed to update the QUOROM Statement (QUality Of Reporting Of Meta-analyses). PRISMA provide an accepted, evidence-based minimum set of items for reporting in systematic reviews, which have been updated to address several conceptual and practical advances in the science of systematic reviews.

Search Strategy and Eligibility Criteria

A systematic search of the literature was conducted to identify studies that investigated SFA as a primary intervention method for people with aphasia. Electronic searches of the following databases were conducted, with the last search in November 2017, using the EBSCOHOST platform: Academic Search Complete, CINAHL Plus with Full Text, E-Journals, MEDLINE with Full Text, PsycINFO, ERIC and the Aphasia Treatment website of the Academy of Neurologic Communication Disorders (http://aphasiatx.arizona.edu/).

The search strategy comprised the following terms:

1. Semantic feature analysis

2. Semantic cues

3. 1 or 2

4. Aphasia

5. Dysphasia

6. 4 or 5

7. Naming

*8. Word finding difficult**

9. 7 or 8

10. 6 or 9

*11. Therap**

*12. Treat**

13. Intervention

14. 11 or 12 or 13

15. 3 and 10 and 14.

After removal of duplicate studies, material resulting from the searches was screened against the eligibility criteria. Studies were considered eligible if they were research reports and were published in English. Studies that combined SFA with other treatment approaches were excluded, when it was impossible to delineate specifically the effects of SFA. Where eligibility could not be assessed on the basis of the title and abstract alone, the full text was obtained.

Study selection: Screening and data extraction

We found 357 abstracts that mentioned Semantic Feature Analysis (SFA) in their abstract and 1500 abstracts that mentioned “semantic cues”. Of these, 145 ~~136~~ were relevant to aphasia / dysphasia and 130 addressed “naming” and / or “word finding difficult*”. Of these, 54 were considered for this review as they also mentioned therapy / treatment / intervention. The full text was obtained for these 54 articles. Of these, seven were excluded as they used different therapy methods, like cueing hierarchy approach (Linebaugh, Shisler, & Lehner, 2005), multi cue computer program (Doesborgh et al., 2004; van Mourik, Verschaeve, Boon, Paquier, & van Harskamp, 1992), personal cueing in

natural settings (Olsen, Freed, & Marshall, 2012), phonological components analysis (PCA) (Leonard, Rochon, & Laird, 2008), orthographic cueing (Leonard, Rochon, & Laird, 2004) and a different semantic approach which compared a phonological and orthographic approach (Lorenz & Ziegler, 2009). One study was excluded as it evaluated SFA in participants with primary progressive aphasia and Alzheimer's Disease (Hung et al., 2017). In this review, only studies that used SFA treatment with semantic feature *generation* have been included. Reports on treatment studies involving semantic feature review or verification have been excluded. Thus 15 articles were excluded, as they reported on a different semantic features approach, such as semantic feature verification rather than generation, or combined SFA with other treatment approaches in the same therapy protocol, such as response elaboration training (RET), communication based therapy, semantic priming, semantic judgment tasks, auditory concept feature and gesturing treatment (Antonucci, 2014a; Boo & Rose, 2011; Cameron, Wambaugh, Wright, & Nessler, 2006; Carragher, Conroy, Sage, & Wilkinson, 2012; Conley & Coelho, 2003; Edmonds & Kiran, 2006; Hashimoto, 2016; Kintz, Wright, & Fergadiotis, 2016; Kiran & Roberts, 2010; Knoph, Simonsen & Lind, 2017; Law, Wong, Sung, & Hon, 2006; Lowell, Beeson, & Holland, 1995; Raymer, Rodriguez, & Rothi, 2007; Wallace & Kimelman, 2013; Wambaugh, Mauszycki, Cameron, Wright, & Nessler, 2013). Moreover, one study was excluded as it evaluated treatment integrity of elaborated SFA (Kladouchou, Papathanasiou, Efstratiadou, Christaki & Hilari, 2017). An additional seven studies were excluded, as they were not research reports (Antonucci, 2014b; Bose & Buchman, 2007; Boyle, 2010; Durand & Ansaldo, 2014; Kiran & Bassetto, 2008; Maddy et al., 2014; van Hees, Mcmahon, Angwin, De Zubicaray, & Copland, 2014a). Lastly, two studies were excluded because they were not relevant to naming, instead one was treating oral reading (Kiran & Viswanathan, 2008), and the other comprehension SFA (Munro & Siyambalapitiy, 2016). The remaining 21 articles were included in the review. The selection process of the articles is illustrated in figure 2.

The 21 studies covered six main areas: confrontation naming of nouns studies, confrontation naming of verbs studies, studies covering both nouns and verbs, connected speech – discourse studies of which two were group studies, multilingual study, and studies where SFA was compared with other approaches, like Phonological Components Analysis (PCA) (Hashimoto, 2012; Neumann, 2017; Sadeghi, Baharloei, Zadeh, & Ghasisin, 2017; van Hees, Angwin, McMahon, & Copland, 2013).

[figure 2 about here]

Critical Appraisal and Methodological Quality

We appraised the methodological quality of included studies and assigned levels of evidence as an indication of risk of bias. Two aphasia-specialist speech-language pathologists critically evaluated the included studies for their methodological quality. All studies were single case studies (N=21). The Single Case Experimental Design Scale (SCEDS) critical appraisal tool (Tate et al., 2008) was used to examine the quality of the studies. SCEDS is an 11-point scale evaluating the methodological quality of experimental single case studies. A perfectly designed and executed study would receive a summative score of 11 across eleven different criteria. A score of 1, per criterion, is given if the study adequately addresses the specified quality item and a score of 0 is given if the item is poorly addressed or not addressed at all. The eleven specified quality items are: (i) clinical history, (ii) target behaviors, (iii) design, (iv) baseline, (v) sampling behavior during treatment, (vi) raw data record, (vii) inter-rater reliability, (viii) independence of assessors, (ix) statistical analysis, (x) replication and (xi) generalization. All included studies were evaluated with SCEDS by two raters. When disagreements between raters were present, an average score was calculated. The first author randomly selected six studies (29%) and re-calculated SCEDS scores to determine intra-rater reliability. Intra-rater reliability was ICC=1.0 (100% agreement). To reduce bias and ensure ratings were not dependent upon one another, re-scoring was completed two weeks after the initial scoring.

Level of evidence was also assigned to each of the studies. Level of evidence refers to the hierarchy of study designs based on the ability of the design to protect against bias. While there is no one universally accepted hierarchy, randomized controlled trials (RCTs) are considered to be the design least susceptible to bias, and various hierarchies follow from there through observational studies and non – experimental designs. We followed the Scottish Intercollegiate Guidelines Network (2010) hierarchy, where RCTs, systematic reviews of RCTs and meta-analyses are considered level 1 evidence; case control and cohort studies and systematic reviews of these are considered level 2 evidence; non-analytic studies, such as case reports and case series are considered level 3 evidence; and expert opinion is considered level 4 evidence. Full information on this classification system is available on http://www.sign.ac.uk/assets/sign118.pdf.

Phase of treatment was also considered for each study, using the coding of Robey and Schultz (1998 & 2004), which is a five – phase model: Phase 1 studies are pre – efficacy studies, where the goal is to determine if there is evidence to suggest that the treatment has therapeutic value. Phase 2 are pre- efficacy studies, where the goal is to develop, standardize, validate, and optimize procedures to explain why a therapy works and who are the ideal candidates. Phase 3 are efficacy studies, where treatment is tested for efficacy under ideal conditions. Phase 4 are effectiveness studies, where treatment is tested for effectiveness under ordinary conditions of use. Lastly, phase 5 are effectiveness studies exploring efficiency, cost-benefit, and patient reported outcomes such as satisfaction and quality of life.

Treatment outcomes and clinical efficacy

As well as describing the treatment outcomes of included studies, the clinical efficacy of SFA was determined by calculating effect sizes. Effect sizes could be calculated only in those studies that reported sufficient data. To calculate, it was necessary to determine the individual values for the pre- treatment and post-treatment phases for each set of trained items. *Cohen's d* statistic was used to calculate effect size as described by Busk and Serlin (1992). The magnitude of change in performance was determined according to the benchmarks for lexical retrieval studies described by Beeson and Robey (2006). The benchmarks were 4.0, 7.0, and 10.1 for small, medium, and large effect sizes respectively.

Where *Cohen's d* could not be calculated, the percent of non-overlapping data (PND) was calculated. PND is the most widely used method of calculating effect size in single case experimental designs (Gast, 2010; Schlosser, Lee, & Wendt, 2008). PND is the percentage of phase B data points (the treatment phase) that do not overlap with phase A data points (baseline or no treatment). To determine the magnitude of effect, benchmarks put forth by Scruggs et al. (1987) were used. PND scores higher than 90% were considered to demonstrate a highly effective treatment, PND of 70–90% were interpreted as a moderate treatment outcome and PND scores of 50–70% were considered a questionable effect. PND scores less than 50% were interpreted as an ineffective intervention since performance during intervention had not affected behavior beyond baseline performance.

Results

Study selection

Twenty-one studies were included in this systematic review. The studies cover six different research areas. Nine studies investigated SFA with confrontation naming of nouns (Boyle, 2004; Boyle & Coelho 1995; Coelho, McHugh & Boyle, 2000; Davis & Stanton, 2005; DeLong, Nessler, Wright, & Wambaugh, 2015; Hashimoto & Frome, 2011; Massaro & Tompkins, 1994; Mehta & Isaki, 2016; Rider, Wright, Marshall, & Page, 2008). Two studies examined SFA with confrontation naming of verbs (Wambaugh & Ferguson, 2007; Wambaugh, Mauszycki, &Wright, 2014) and a further two tested SFA with confrontation naming of nouns and verbs (Kristensson, Behrns, & Saldert, 2015; Marcotte & Ansaldo, 2010). Kristensson's study additionally explored everyday conversation and functional communication outcomes. Connected speech – discourse - was examined in one study (Peach & Reuter, 2010), group SFA was evaluated in two studies (Antonucci, 2009; Falconer & Antonucci, 2012), and multilingual SFA was tested in one study (Knoph, Lind, & Simonsen, 2015). Finally, four studies compared SFA with other approaches, like Phonological Components Analysis (PCA) (Hashimoto, 2012; Neumann, 2017; Sadeghi, Baharloei, Zadeh, & Ghasisin, 2017; van Hees et al., 2013). Before presenting the characteristics and details of the above studies their methodological quality will be considered.

Critical Appraisal and Methodological Quality

Across the 21 studies, scores on the SCEDS ranged from 8.0 to 11 with an average score of 9.55 out of 11 (Table 1). After SCEDS scoring, level of evidence was assigned for the studies. All studies were determined to be well – designed non – experimental / non – analytic studies and assigned a level 3 rating, except of Marcotte and Ansaldo (2010), which was classified as an observational controlled study.

[table 1 about here]

Phase of treatment was obtained for all studies. Chronologically earlier studies, from 1994 until 2007 and Hashimoto's and Frome's study (2011), were Phase 1 studies (see Table 1),

i.e., pre–efficacy studies (n=11), where the goal was to determine if there was evidence to suggest that the treatment had therapeutic value. All other studies, except for Rider et al., (2008) were Phase 2 pre-efficacy studies (n=9), where the goal was to develop, standardize, validate, and optimize procedures to explain why SFA worked and who were the ideal candidates. Rider and colleagues' study (2008) was a Phase 3 efficacy study, where treatment was tested for efficacy under ideal conditions. The prevalence of high SCEDS scores suggests the included studies were of good/adequate methodological quality, despite being pre-efficacy studies.

Characteristics of studies:

Type and duration of treatment

Study and participant characteristics are shown in Tables 2 and 3. Table 2 details the number of participants, type of SFA treatment, dosage and duration of treatment and total amount of treatment expressed in minutes. A total of 55 participants have been treated in the included studies. Total amount of treatment ranged from 315 minutes (Sadeghi et al., 2017) to 1500 minutes (Boyle, 2004) [mean (SD) = 1019.69 (337.17)].

SFA of nouns

Nine studies, with a total of 18 monolingual individuals, tested SFA in confrontation naming tasks of single nouns (Boyle, 2004; Boyle & Coelho 1995; Coelho et al., 2000; Davis & Stanton, 2005; DeLong et al., 2015; Hashimoto & Frome, 2011; Massaro & Tompkins, 1994; Mehta & Isaki, 2016; Rider et al., 2008). Treatment duration ranged from five to 12 weeks and treatment was delivered in two to three 60 minute sessions per week, with a total amount of treatment of 12 to 24 hours [mean (SD)= 18 (4.38)].

SFA of verbs

Two studies, with five monolingual participants, applied SFA in confrontation naming tasks that targeted single verbs (Wambaugh & Ferguson, 2007; Wambaugh et al., 2014). The treatment duration was four weeks and treatment was delivered in three 45 - 60 minutes' sessions per week.

SFA of nouns and verbs

Two SFA studies combined confrontation naming tasks of single nouns and verbs (Kristensson et al., 2015; Marcotte & Ansaldo, 2010). In Marcotte and Ansaldo's (2010) study the treatment duration for the individual was three weeks and he had three 60 minutes' sessions per week resulting in nine hours of therapy in total. In Kristensson and

colleagues' (2015) study the three participants received 20 hours of treatment delivered in 20 sessions lasting 60 minutes each for a period of five to six weeks.

Discourse SFA

Discourse SFA was evaluated in three studies, one using an individual approach (Peach & Reuter, 2010) and two using a group approach (Antonucci, 2009; Falconer & Antonucci, 2012). Individual discourse SFA was evaluated with two participants, one monolingual and one bilingual (Peach & Reuter, 2010). Treatment was delivered in 50 minutes' sessions and lasted ten weeks, with a total amount of treatment of 11-12 hours. Group approach SFA was tested in two studies (Antonucci, 2009; Falconer & Antonucci, 2012), with seven monolingual participants, for seven weeks, with a small difference on the amount of hours in each study. In Antonucci (2009) each session ranged from 60 to 90 minutes and in Falconer and Antonucci (2012) from 90 - 120 minutes, resulting in a total amount of treatment of 1050 - 1470 minutes [mean (SD) = 1260 (296.98)].

Multilingual SFA

Multilingual SFA was tested in one study (Knoph, Lind, & Simonsen, 2015), with one quadrilingual participant, for two and a half weeks, each session ranged from 45 to 55 minutes, resulting in a total amount of 1320 minutes.

Comparing SFA to PCA

Four studies compared SFA with PCA (Hashimoto, 2012; Neumann, 2017; Sadeghi et al., 2017; van Hees et al., 2013) in a total of 18 participants. In the Hashimoto (2012) study, two participants were seen twice weekly and had two 45-60 minute sessions on each of these two days for 15 to 25 weeks. In the van Hees et al (2013) study, eight participants received three 45-90 minute sessions per week for four weeks. In the Neumann (2017) study, four participants were seen two to three times per week for a two-hour session, the duration of treatment varied from two to six and a half weeks. In Sadeghi et al (2017), four participants received seven 45 minute sessions for two weeks.

Participant characteristics

Table 3 presents the demographic characteristics of the 55 participants from the 21 reviewed studies. Considerable heterogeneity was found across the participants in terms of age and time post onset. Age ranged from 24 to 80 years, with a mean (SD) age of 55.39 (12.66). Time post onset ranged from 4 to 384 months, with a mean (SD) of 59.75

(68.70) months. Thirty-one participants were men and 24 were women. Of the participants, 21 were described as non–fluent and 33 as fluent (one was not reported). Aphasia was due to a stroke in 51 individuals and to traumatic brain injury in four individuals (neuropathology for three individuals was not reported). Aphasia severity was reported or derived from the aphasia quotient (AQ) of the WAB in 14 studies. Four studies based aphasia severity on a different test and three did not report severity. One participant presented with very severe aphasia, four with severe, three with moderate to severe, 23 with moderate, three with mild to moderate, and 14 with mild aphasia. Aphasia type was not reported for six participants. Of the remaining, 14 had Broca's aphasia, 16 anomic, five Wernicke's, nine conduction, one global, one mixed and three transcortical motor aphasia.

[table 2 about here]

[table 3 about here]

Synthesis of results

Treatment outcomes

The main treatment outcomes of the reviewed studies are summarized in Table 4. Improvement in naming of trained items was found for 45 participants (81.82%). Maintenance of naming of the trained items was reported for 32 participants (58.18%). Generalization effects ranged from negligible (e.g., Rider et al., 2008) to strong (Boyle, 2004). The percentage of generalization to untrained items for all studies was small (40%).

[table 4 about here]

In relation to aphasia type and the outcome of SFA therapy, we looked firstly at improvement on the trained items. Twelve of the 14 (85.71%) participants with Broca's aphasia, 13 of the 16 anomic participants (81.25 %), four of the five (80%) individuals with Wernicke's aphasia, and all nine with conduction aphasia and three with transcortical motor aphasia (100%) showed improvement on naming of trained items. Negative outcomes were

found for the two participants with global and mixed aphasia. In terms of maintenance, the findings were positive for eight (50%) of the anomic participants, seven (50%) participants with Broca's aphasia and all those with conduction and transcortical motor aphasia (100%), whereas only two (40%) of participants with Wernicke's aphasia, and none of the two individuals with global or mixed aphasia showed a maintenance effect. In terms of generalization to untreated items, it was mostly the individuals with Broca's aphasia that showed positive gains (57.14 %). All other aphasia type participants showed minimal gains on generalization to untreated items. Specifically, gains were reported for 33.33 % of the participants with conduction aphasia, 25% of Wernicke's aphasia, 16.67 % of those with anomic aphasia and 37.5% of the individuals with transcortical motor aphasia.

All studies assessed post - therapy gains immediately after treatment ended. The number of assessments and the timing of follow-up assessments varied (table 5). Overall, three studies assessed gains only once post-therapy (Knoph, Lind, & Simonsen, 2015; Marcotte & Ansaldo, 2010; Sadeghi et al., 2017) and 18 included follow-up/maintenance assessments. The majority of the studies (n=12) assessed maintenance up to six weeks after the end of treatment. Five studies assessed maintenance up to 2-4.5 months after the end of treatment (Boyle & Coelho, 1995; Coelho et al., 2000; Kristensson et al., 2015; Mehta & Isaki, 2016; Peach & Reuter, 2010). Only one study went beyond 4.5 months and had multiple follow-up assessments up to a year (Davis and Stanton, 2005).

[table 5 about here]

Clinical efficacy

Effect sizes for treatment outcomes were reported in eleven studies (Antonucci, 2009; DeLong et al.,2015; Falconer & Antonucci, 2012; Hashimoto & Frome, 2011; Hashimoto, 2012; Knoph et al., 2015; Kristensson et al., 2015; Peach & Reuter, 2010; Rider et al.,2008; van Hees et al.,2013; Wambaugh et al., 2014;). Calculation could not be performed for three studies (Davis & Stanton, 2005; Marcotte & Ansaldo, 2010; Mehta & Isaki, 2016).

The first author of the review calculated effect sizes for two studies (n = 2) (Boyle, 2004; Wambaugh & Ferguson; 2007), as well as average effect sizes for eight studies (n = 24) (Antonucci, 2009; DeLong et al., 2015; Hashimoto & Frome, 2011; Kristensson et al., 2015;

Neumann, 2017; Rider et al., 2008; Sadeghi et al., 2017; Wambaugh et al., 2014) (Table 6). Average effect sizes were calculated when data were collected and reported on two or more trials at one-time point. Further effect sizes were calculated in three studies (n = 4) (Boyle & Coelho, 1995; Coelho et al., 2000; Massaro & Tompkins, 1994) based on substitute data from other phases, following the recommendation of Beeson and Robey (2006). Large effect sizes were present for eight participants (d = 10.07 – 19.23). Medium effect sizes were present for six participants (d = 7.00 – 9.58). Small effect sizes were present for eleven participants (d = 4.14 – 6.89). For 17 participants, effect sizes were negligible and for five there was no change. Effect size and PND could not be calculated for five participants from the studies of Marcotte and Ansaldo (2010), Mehta and Isaki (2016) and one participant from DeLong et al. (2015) and Antonucci (2009) studies.

PND was calculated for three studies (Boyle, 2004; Davis & Stanton, 2005; Peach & Reuter, 2010), for three participants for whom effect sizes could not be calculated. A large treatment effect (PND > 90%) was evident for two participants and a moderate treatment effect for one participant (PND = 85%). When examining clinical efficacy using PND, treatment was highly effective for the majority of participants. None of the participants had PND scores consistent with ineffective treatment.

[table 6 about here]

Discussion

The purpose of this review was to evaluate the quality of SFA therapy studies in aphasia; detail their characteristics and synthesize their results. We reviewed 21 studies reporting on 55 persons with aphasia. Improvement in naming of trained items was found for 45 participants (81.82%). Thus, SFA improved treated items for the majority of participants. Yet, effect size calculations indicated that there was a small or less than small treatment effect for a substantial proportion of participants (28/45, 62.22%). Moreover, although findings suggest that treatment was effective for improving naming of trained items, limited generalization to untrained items and connected speech was reported (40%).

Maintenance of the trained items post therapy was reported for 32 participants (58.18%). Maintenance of therapy gains can be affected by factors like the timing of assessment, treatment dosage and duration (Boyle, 2010). Timing of assessment for maintenance effects varied (see Table 5). This variation may affect results, as when the evaluation is closer to the end of the intervention, maintenance of gains is more likely than when maintenance is assessed after a longer period. Looking at short-term maintenance, from the 21 studies, short - term post – therapy gains (two weeks) were reported in only five studies (DeLong et al., 2015; Massaro & Tompkins, 1994; Wambaugh et al., 2014; Wambaugh & Ferguson; 2007; van Hees et al., 2013). Eleven of the 20 participants (55%) in these studies showed a maintenance effect. If we consider longer-term post – therapy gains, six studies looked at two months or more post therapy, with 5 of 10 participants (50 %) showing maintenance of treatment gains (Boyle & Coelho, 1995; Coelho et al., 2000; Davis & Stanton, 2005; Kristensson et al., 2015; Mehta & Isaki, 2016; Peach & Reuter, 2010). Though the results seem to confirm that closer to the end of therapy gains are more likely to be maintained, we need to interpret this with caution as the number of participants assessed in the longer term ($\geq$ 2 months) is small.

Results of generalization to untreated items ranged from strong (e.g., Boyle, 2004) to negligible (e.g., Rider et al., 2008; Wambaugh et al., 2014; Wambaugh & Ferguson, 2007). Positive generalization outcomes were evident for 40% of participants. It is argued that generalization may be related to the underlying mechanism of how SFA works. That is, if SFA has a semantic network repair function, then untreated items that belong to the same semantic category as trained items will indirectly benefit from treatment. Items that lie outside of the semantic network would not be likely to benefit. However, if SFA functions as a self-employed ‘‘semantic cueing strategy’’, as Lowell and colleagues (1995) suggested, it would be expected that semantically related and unrelated items would improve when the strategy is implemented successfully. In this review, it has not been possible to evaluate this hypothesis as limited information was provided in most studies on the nature of generalization. However, Boyle (2004) performed a post hoc analysis of categorical membership of treated and untreated experimental stimuli and found that generalization occurred to untreated items that were not members of the same categories as treated items. Generalization to unrelated items suggested that SFA functioned as a mediating strategy for naming those items.

One study reported on a multilingual participant (Knoph et al., 2015) and found naming improvement in the untreated languages. Similar findings have been reported in prior studies where semantic feature verification has been used with bilingual speakers (Edmonds & Kiran, 2006; Kiran & Roberts, 2010), with cross-linguistic transfer in some conditions for some participants. It has been suggested that cross-linguistic transfer is difficult to achieve (Ansaldo & Ghazi Saidi, 2014; and Faroqi - Shah et al., 2010). Yet, Knoph and colleagues (2015) hypothesized that the semantic nature of SFA therapy would lead to cross -linguistic transfer, and their results partly supported their hypothesis.

Although all studies focus on treating word finding difficulties in aphasia, pulling their results together is challenging due to the expected heterogeneity of various study components. A variety of aphasia types has been evaluated. Individuals with Broca's, Wernicke's, anomic, conduction, global, and transcortical motor aphasia syndromes have been included. Dividing participants to the broad categories of fluent and non – fluent aphasia, people with fluent aphasia are the most represented subtype in the reviewed studies (33/55, 60%). In terms of aphasia severity, the main body of the participants (72.73%) had mild (n=14), mild-moderate (n=3), or moderate (n=23) aphasia. Overall, results suggested that SFA as a treatment for word finding difficulties may be more effective for persons with fluent and moderate or mild aphasia (Antonucci, 2009; Boyle, 2004; Coelho et al., 2000; Hashimoto, 2012) compared to those with non – fluent and more severe aphasia (Hashimoto & Frome, 2011; Kristensson et al., 2015). However, Boyle (2010) in a review of SFA treatments for nouns found that participants with severe aphasia also had positive responses. Lowell et al. (1995) suggested that aphasia severity and poor non-verbal cognitive skills were determining factors for participants who did not show improvement post therapy. Wambaugh and colleagues (2013) also suggested that different profiles of language, memory, and cognition might be associated with different responses to SFA. Further research with large numbers of participants is necessary in order to begin to unravel the impact of different aphasic profiles and severities on the efficacy of SFA.

Another important consideration is that treatments, which are called SFA, are not always the same in terms of their treatment protocols. Many studies changed the traditional SFA protocol in various ways, such as modifications to the semantic feature categories (Mehta &

Isaki, 2016; Wambaugh et al., 2014; Wambaugh & Ferguson, 2007), eliciting fewer features (Hashimoto & Frome, 2011; Mehta & Isaki, 2016), writing the features in addition to or instead of saying them (Hashimoto & Frome, 2011), following different treatment stages (Davis & Stanton, 2005), and adding new factors, such as independent homework (Falconer & Antonucci, 2012). This variability again makes it difficult to determine which aspects of SFA were most effective.

Different treatment outcomes could also be due to different treatment durations, dosages and total amount of treatment. Therefore, another limiting factor is the lack of a standardized dosage and treatment duration across studies. Some studies, like Hashimoto and Frome (2011) reported longer treatment sessions over a shorter duration. Across the studies reviewed, duration of treatment varied from two weeks to twelve weeks [mean (SD) = 5.92 (2.56)]. Treatment sessions per week also varied from two to four sessions [mean (SD) = 2.64 (0.59)], and duration of sessions varied from 45 minutes to 90-120 minutes [mean (SD) = 63.28 (18.42)]. The most common duration per session was one hour (identified in eight different studies). It may be that total amount of treatment may relate to treatment outcomes. The findings of this review partly support this finding. There were eight studies with low amount of treatment, i.e. 315-720 minutes (Davis & Stanton, 2005; Marcotte & Ansaldo, 2010; Mehta & Isaki, 2016; Peach & Reuter, 2010; Rider et al., 2008; Sadeghi et al., 2017; Wambaugh et al., 2007 & 2014). Eighteen of the 19 participants in these studies made gains in naming post-therapy, nine of the 19 maintained these gains and seven generalized to untreated items. In the six studies with high overall treatment amount (1260-1470 minutes), 11 of 11 participants made gains post-therapy, and 9 of 10 maintained these gains and generalized to untreated items (Boyle, 2004; Coelho et al., 2000; Falconer & Antonucci, 2012; Hashimoto & Frome, 2011; Hashimoto, 2012).

Despite the complicating factors of variability of treatment procedures, dosage, duration and changes to the traditional SFA protocol, this systematic review of SFA studies suggests that SFA is an effective intervention that can elicit positive therapy outcomes. Synthesizing the findings of 21 single case and case series studies suggests that SFA is effective in improving treated items and has a small effect on generalization to untrained items. In summary, the evidence-base for SFA as a therapeutic intervention is growing, but further research with larger numbers of participants is warranted to examine differential gains across aphasia types

and explore generalization to untreated items and longer term maintenance with greater confidence.

Acknowledgments

This research was supported by a doctoral studentship to the first author by City, University of London School of Health Sciences.

References

References marked with an asterisk (*) indicate studies included in the review.

*Antonucci, S.M. (2009). Use of semantic feature analysis in group aphasia treatment. Aphasiology, 23 (7–8),854–866.

Antonucci, S. M. (2014a). How do persons with aphasia use semantic feature information during lexical retrieval: Evidence from verbal-description and naming-to-definition of living and non - living concepts. *Journal of the International Neuropsychological Society*,20 (S1).

Antonucci, S. M. (2014b). What matters in semantic feature processing for persons with stroke-aphasia: Evidence from an auditory concept–feature verification task. *Aphasiology*,*28 (7)*, 823–839.

Aphasia Treatment Evidence Tables. (2015). Academy of Neurologic Communication Disorders in profile 2015. Retrieved from http://aphasiatx.arizona.edu/.

Beeson, P., & Robey, R. (2006). Evaluating single-subject treatment research: lessons learned from the aphasia literature. *Neuropsychological Review, 16*, 161–169.

Best, W. & Nickels, L. (2000). From theory to therapy in aphasia: Where are we now and where to next?. *Neuropsychological Rehabilitation*, *10*(3), 231-247

Boo, M., & Rose, M. L. (2011). The efficacy of repetition, semantic, and gesture treatments for verb retrieval and use in Broca's aphasia. *Aphasiology, 25(2)*, 154–175.

Bose, A. and Buchanan, L. (2007). A cognitive and psycholinguistic investigation of neologisms. *Aphasiology, 21(6-8*), 726-738.

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

*Boyle, M. (2004). Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech- Language Pathology*, *13*, 236–249.

Boyle, M. (2010). Semantic feature analysis treatment for aphasic word retrieval impairments: what's in a name? *Top Stroke Rehabilitation, 17*, 411–422.

Busk, P.L., & Serlin, R.C. (1992). Meta-analysis for single-case research. *Single-case research design and analysis*. In: Kratochwill TR, Levin JR, editors. Hillsdale, NJ: Lawrence Erlbaum, p. 187–212.

Byng, S. (1988). Sentence processing deficits: Theory and therapy. *Cognitive Neuropsychology*, *5*(6), 629-676.

Cameron, R., Wambaugh, J., Wright, S., & Nessler, C. (2006). Effects of a combined semantic/phonologic cueing treatment on word retrieval in discourse. *Aphasiology*,*20*(2-4), 269-285.

Carragher, M., Conroy, P., Sage, K., & Wilkinson, R. (2012). Can impairment-focused therapy change the everyday conversations of people with aphasia? A review of the literature and future directions. *Aphasiology, 26(7),* 895–916.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407–428.

*Coelho, C., McHugh, R., & Boyle, M. (2000). Semantic feature analysis as a treatment for aphasic dysnomia: a replication. *Aphasiology, 14*, 133–142.

Conley, A., & Coelho, C.A. (2003). Treatment of word retrieval impairment in chronic Broca's aphasia. *Aphasiology, 17*, 203–211.

Davis, G. A. (2000). *Aphasiology: Disorders and clinical practice*. Boston: Allyn & Bacon.

*Davis L.A., & Stanton, S.T. (2005). Semantic Feature Analysis as a Functional Therapy Tool. *Contemporary Issues in Communication Science and Disorders, 32*, 85–92.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*, 801–838.

Dell, G. S., Lawler, E. N., Harris, H. D., & Gordon, J. K. (2004). Models of errors of omission in aphasic naming. *Cognitive Neuropsychology*, *21*, 125–145.

*DeLong, C.., Nessler, C., Wright, S., & Wambaugh, J. (2015). Semantic Feature Analysis: Further Examination of Outcomes. *American Journal of Speech-Language Pathology*, 24 (4), 864 -879.

Doesborgh, S. J., Van De Sandt-Koenderman, M. W., Dippel, D. W., Van Harskamp, F., Koudstaal, P. J., & Visch-Brink, E. G. (2004). Effects of semantic treatment on verbal communication and linguistic processing in aphasia after stroke: A randomized controlled trial. *Stroke*, *35*, 141–146.

Drew, R. & Thompson, C. (1999). Model-Based Semantic Treatment for Naming Deficits in Aphasia. *Journal of Speech Language And Hearing Research*, *42*(4), 972 - 989.

Durand, E. & Ansaldo, A.I. (2014) Recovery from Anomia Following Semantic Feature Analysis: Therapy-induced Neuroplasticity Relies upon a Circuit Involving Language and Motor Processing Areas. The Mental Lexicon. Special Issue: *The Neural Correlates of Lexical Processing, 21,* 195-215.

Eales, C. & Pring, T. (1998). Using individual and group therapy to remediate word finding difficulties. *Aphasiology, 12*(10), 913-918.

Edmonds, L.A., & Kiran, S. (2006). Effect of semantic naming treatment on cross linguistic generalization in bilingual aphasia. *Journal of Speech, Language, and Hearing Research., 49*, 729–748.

*Falconer, C., & Antonucci, S.M. (2012). Use of semantic feature analysis in group discourse treatment for aphasia: Extension and expansion. *Aphasiology, 26 (1),* 64–82. 20.

Fotiadou D., Northcott S., Chatzidaki A., & Hilari K. (2014) Aphasia blog talk: How does stroke and aphasia affect a person's social relationships? Aphasiology, 28, 1281-1300

Franklin, S., Buerk, F., & Howard, D. (2002). Generalised improvement in speech production for a subject with reproduction conduction aphasia. *Aphasiology, 16*(10-11), 1087-1114.

Gast, D. (2010). *Single subject research methodology in behavioural sciences.* New York: Routledge.

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes, 21,* 817–855.

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *The Boston Diagnostic Aphasia Examination: BDAE-3 Short Form.* Philadelphia, PA: Lippincott, Williams, and Wilkins.

Goodglass, H., & Wingfield, A. (1997). *Anomia: Neuroanatomical and cognitive correlates*. San Diego, CA: Academic Press.

Greenwood, A., Grassly, J., Hickin, J., & Best, W. (2010). Phonological and orthographic cueing therapy: A case of generalised improvement. *Aphasiology, 24*(9), 991-1016.

Haarbauer-Krupa, J., Moser, L., Smith, G., Sullivan, D.M., & Szekeres, S.F. (1985a). Cognitive rehabilitation therapy: middle stages of recovery. In: Ylvisaker M, ed. *Head Injury Rehabilitation: Children and Adolescents.* San Diego, CA: College-Hill Press, 287–310.

Haarbauer-Krupa, J., Moser, L., Smith, G., Sullivan, D.M., & Szekeres, S.F. (1985b). Cognitive rehabilitation therapy: late stages of recovery. In: Ylvisaker M, ed. *Head Injury Rehabilitation: Children and Adolescents.* San Diego, CA: College-Hill Press, 311–340.

*Hashimoto, N., & Frome, A. (2011). The use of a modified semantic features analysis approach in aphasia. *Journal of Communication Disorders, 44,* 459–469.

*Hashimoto, N. (2012). The use of semantic-and phonological-based feature approaches to treat naming deficits in aphasia, *Clinical Linguistics & Phonetics, 26*, 518-533.

Hashimoto, N. (2016). The role of effects of thematic relations on picture naming abilities across the lifespan. *Aging, Neuropsychology, & Cognition, 23,* 499-512.

Helm-Estabrooks, N. (1992). *Aphasia Diagnostic Profiles.* Chicago: Riverside.

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.

Hilari, K., Needle, J. J., & Harrison, K. L. (2012). What are the important factors in health-related quality of life for people with aphasia? A systematic review. *Archives of Physical Medicine and Rehabilitation, 93,* 86-95.

Hough, M. (1993). Treatment of Wernicke's aphasia with jargon: A case study. *Journal Of Communication Disorders, 26*(2), 101-111.

Howard, D. (2000). Cognitive neuropsychology and aphasia therapy: the case of word retrieval. In: Papathanasiou I, ed. *Acquired Neurogenic Communication Disorders: A Clinical Perspective.* London, England: Whurr, 76–99.

Howard, D. (1994). The Treatment of Acquired Aphasia. *Philosophical Transactions Of The Royal Society B: Biological Sciences, 346*(1315), 113-120.

Howard, D., & Gatehouse, C. (2006). Distinguishing semantic and lexical word retrieval deficits in people with aphasia. *Aphasiology, 20,* 921–950.

Howard, D., Patterson, K., Franklin, S., Orchard-lisle, V., & Morton, J. (1985). The facilitation of picture naming in aphasia. *Cognitive Neuropsychology, 2*(1), 49-80.

Hung, J., Bauer, A., Grossman, M., Hamilton, R., Coslett, H., & Reilly, J. (2017). Semantic Feature Training in Combination with Transcranial Direct Current Stimulation (tDCS) for Progressive Anomia. *Frontiers in Human Neuroscience, 11*.

Kertesz, A. (2007). *Western Aphasia Battery—Revised.* San Antonio, TX: Harcourt Assessment.

Kintz, S., Wright, H., & Fergadiotis, G. (2016). Semantic knowledge use in discourse produced by individuals with anomic aphasia. *Aphasiology, 30* (9), 1012-1025.

Kiran, S., & Bassetto, G. (2008). Evaluating the Effectiveness of Semantic-Based Treatment for Naming Deficits in Aphasia: What Works? *Seminars in Speech and Language, 29 (1),* 71-82.

Kiran, S., & Viswanathan, M. (2008). Effect of model based treatment on oral reading abilities in severe alexia: A case study. *Journal of Medical Speech Pathology,16 (1),* 43-59.

Kiran, S., & Roberts, P.M. (2010). Semantic feature analysis treatment in Spanish-English and French English bilingual aphasia. *Aphasiology, 24,* 231–261.

Kladouchou, V., Papathanasiou, I., Efstratiadou, E., Christaki, V., & Hilari, K. (2017). Treatment integrity of elaborated semantic feature analysis aphasia therapy delivered in individual and group settings. *International Journal Of Language & Communication Disorders*.

*Knoph, M. I. N., Lind, M., & Simonsen H. G. (2015). Semantic feature analysis targeting verbs in a quadrilingual speaker with aphasia, *Aphasiology, 29 (12),* 1473-1496.

Knoph, M., Simonsen, H., & Lind, M. (2017). Cross-linguistic transfer effects of verb-production therapy in two cases of multilingual aphasia. *Aphasiology, 31*(12), 1482-1509.

*Kristensson, J., Behrns, I., & Saldert, C. (2015). Effects on communication from intensive treatment with semantic feature analysis in aphasia. *Aphasiology, 29 (4),* 466–487.

Laine, M., & Martin, N. (2006). *Anomia: Theoretical and Clinical Aspects. Hove*, UK: Psychology Press.

Law, S.P., Wong, W., Sung, F., & Hon, J. (2006). A study of semantic treatment of three Chinese anomic patients. *Neuropsychological Rehabilitation, 16,* 601–629.

Leonard, C., Rochon, E., & Laird, L. (2004). The benefits of orthographic cueing to naming deficits in aphasia: A case study. *Brain and Language, 19,* 185-186.

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

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences, 3,* 223–232.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–75.

Linebaugh, C.W., Shisler, R.J., & Lehner, L., (2005). Cueing hierarchies and word retrieval: A therapy program. *Aphasiology, 19(1),* 77-92.

Lorenz, A., & Ziegler, W. (2009). Semantic vs. word-form specific techniques in anomia treatment: A multiple single-case study. *Journal of Neurolinguistics, 22,* 515–537.

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

Lowell, S., Beeson, P.M., & Holland, A.L. (1995). The efficacy of a semantic cueing procedure on naming performance of adults with aphasia. *American Journal of Speech-Language Pathology, 4,* 109–114.

Maddy, K.M., Capilouto, G.J., McComas, K.L. (2014). The effectiveness of semantic feature analysis: An evidence-based systematic review. *Annals of Physical and Rehabilitation Medicine, 57,* 254–267.

Maher, L. M., & Raymer, A. M. (2004). Management of anomia. *Topics in Stroke Rehabilitation, 11,* 10–21.

*Marcotte, K., & Ansaldo, AI. (2010). The neural correlates of semantic feature analysis in chronic aphasia: discordant patterns according to the etiology. *Seminars in Speech and Language. Specials Issue on Treatment of Chronic Aphasia: International Perspectives, 31(1),* 52–63.

Marshall, J., Pound, C., White-thomson, M., & Pring, T. (1990). The use of picture/word matching tasks to assist word retrieval in aphasic patients. *Aphasiology, 4*(2), 167-184.

Martin, N., Fink, R., & Laine, M. (2004). Treatment of word retrieval deficits with contextual priming. *Aphasiology, 18*(5-7), 457-471.

*Massaro, M., & Tompkins, C. (1994). Feature analysis for treatment of communication disorders in traumatically brain-injured patients: an efficacy study. *Clinical Aphasiology, 22,* 245–256.

*Mehta S. V., & Isaki E. (2016). A modified Semantic Feature Analysis Approach with two individuals with chronic aphasia. *Contemporary Issues in Communication Science and Disorders,* 43, 129 -138.

Mitchum, C., Haendiges, A., & Berndt, R. (1993). Model-guided treatment to improve written sentence production: A case study. *Aphasiology, 7*(1), 71-109.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151,* 264–269.

Moher, D., Liberati, A., Tetzlaff, J., & Alman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *International Journal of Surgery, 8,* 336-341.

Munro, P., & Siyambalapitiya, S. (2016). Improved word comprehension in Global aphasia using a modified semantic feature analysis treatment. *Clinical Linguistics & Phonetics, 31*(2), 119-136.

*Neumann, Y. (2017). A case series comparison of semantically focused vs. phonologically focused cued naming treatment in aphasia. *Clinical Linguistics & Phonetics,* 1-27.

Nickels, L. (2002). Therapy for naming disorders: revisiting, revising, and reviewing. *Aphasiology, 16,* 935–979.

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

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

Northcott, S., Moss, B., Harrison, K., & Hilari, K. (2016). A systematic review of the impact of stroke on social support and social networks: associated factors and patterns of change. *Clinical Rehabilitation, 30(8),* 811-831.

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.

Paradis, M., Libben, G., & Hummel, K. (1987). *The bilingual aphasia test (English version)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

*Peach, R., & Reuter, K. (200). A discourse-based approach to semantic feature analysis for the treatment of aphasic word retrieval failures. *Aphasiology, 24,* 971–990.

Raymer, A. M., Beeson, P., Holland, A., Kendall, D., Maher, L. M., Martin, N., et al. (2008). Translational research in aphasia: From neuroscience to neurorehabilitation. *Journal of Speech, Language, and Hearing Research, 51(1),* 259–275.

Raymer, A., Thompson, C., Jacobs, B., & Le Grand, H. (1993). Phonological treatment of naming deficits in aphasia: Model-based generalization analysis. *Aphasiology, 7*(1), 27-53.

Renvall, K., Laine, M., & Martin, N. (2007). Treatment of anomia with contextual priming: Exploration of a modified procedure with additional semantic and phonological tasks. *Aphasiology, 21*(5), 499-527.

*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*, 161–172.

Robey, R. R., & Schultz, M. C. (1998). A model for conducting clinical outcome research: An adaptation of the standard protocol for use in aphasiology. *Aphasiology, 12 (9),* 787–810.

Robey, R. R. (2004). A five-phase model for clinical-outcome research. *Journal of Communication Disorders, 37 (5),* 401 - 411.

Robson, J., Pring, T., Marshall, J., Morrison, S., & Chiat, S. (1998). Written communication in undifferentiated jargon aphasia: a therapy study. *International Journal of Language & Communication Disorders*, *33*(3), 305-328.

*Sadeghi Z, Baharloei N, Moddarres Zadeh A, Ghasisin L. (2017). Comparative Effectiveness of Semantic Feature Analysis (SFA) and Phonological Components Analysis (PCA) for Anomia Treatment in Persian Speaking Patients With Aphasia. *Iranian Rehabilitation Journal*, 15 (3), 259-268.

Schlosser, R.W., Lee, D.L., & Wendt, O. (2008). Application of the percentage of non - overlapping data (PND) in systematic reviews and meta-analyses: a systematic review of reporting characteristics. *Evidence Based Communication Assess Intervention, 2,* 163–87.

Schwartz, M.F., & Brecher, A. (2000). A model-driven analysis of severity, response characteristics, and partial recovery in aphasics' picture naming. *Brain and Language, 73(1),* 62–91.

Schwartz, M. F., Dell, G. S., Martin, N., Gabl, S., & Sobel, P. (2006). A case-series of test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language, 54,* 228–264.

Scottish Intercollegiate Guidelines Network. (2010). Management of patients with stroke: Rehabilitation, prevention and management of complications, and discharge planning. A national clinical guideline. Retrieved from http://www.sign.ac.uk/assets/sign118.pdf.

Scruggs, T., Mastropieri, M., & Casto, G. (1987). The quantitative synthesis of single subject research: methodology and validation. *Remedial and Special Education, 8 (2),* 24–33.

Tate, R., McDonald, K., Perdices, M., Tougher, L., Schultz, R., & Savage, S. (2008). Rating the methodological quality of single subject designs and n-of-1 trials: introducing the Single-Case Experimental Design (SCED) Scale. *Neuropsychological Rehabilitation, 18,* 385–401.

*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.

van Hees, S., Mcmahon, K., Angwin, A., De Zubicaray, G., & Copland, D. A. (2014a). Neural activity associated with semantic versus phonological anomia treatments in aphasia. *Brain Language, 129,* 47–57.

van Mourik, M., Verschaeve, M., Boon, P., Paquier, P., & van Harskamp, F., (1992). Cognition in global aphasia: Indicators or therapy. *Aphasiology, 5*, 491-499.

Wambaugh, J. L., Linebaugh, C. W., Doyle, P. J., Martinez, A. L., & Kalinyak-Fliszar, M. M. (2001). Effects of two cueing treatments on lexical retrieval in aphasic speakers with different levels of deficit. *Aphasiology, 15(10/11),* 933–950.

*Wambaugh, J. L., & Ferguson, M. (2007). Application of semantic feature analysis to retrieval of action names in aphasia. *Journal of Rehabilitation Research and Development, 44,* 381–394.

Wambaugh, J.L., Mauszycki, S., Cameron, R., Wright, S., & Nessler, C. (2013). Semantic feature analysis: incorporating typicality treatment and mediating strategy training to promote generalization. *American Journal of Speech-Language Pathology, 22,* 334–369.

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

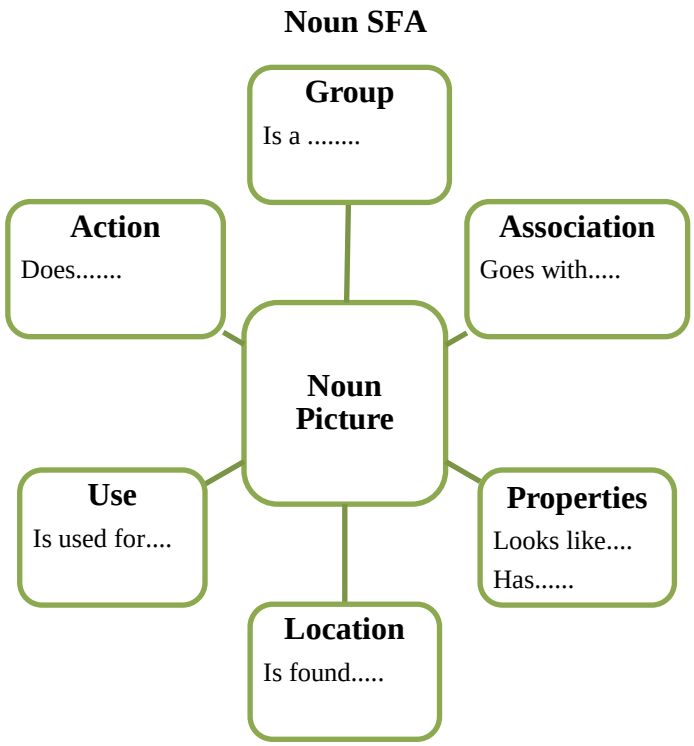
Wallance, S.E., Kimelman, M.D.Z. (2013). Generalization of word retrieval following semantic feature treatment. *NeuroRehabilitation, 32(4),* 899-913.

Wilshire, C. E. (2002). Where do aphasic phonological errors come from? Evidence from phoneme movement errors in picture naming. *Aphasiology, 16*, 169–197.

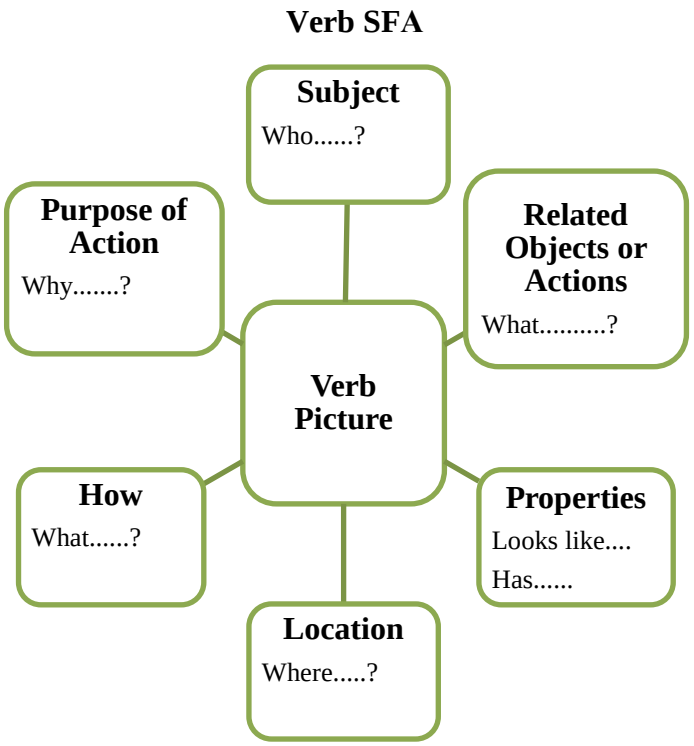
Wisenburn, B., & Mahoney, K. (2009). A meta-analysis of word-finding treatments for aphasia. *Aphasiology*, *23*(11), 1338-1352.

Ylvisaker, M., & Szekeres, S. (1985). *Cognitive-language intervention with brain injured adolescents and adults.* Paper presented at: Annual Convention of the Illinois Speech-Language-Hearing Association. Chicago, Illinois.

Figure 1: The feature analysis charts for nouns and verbs

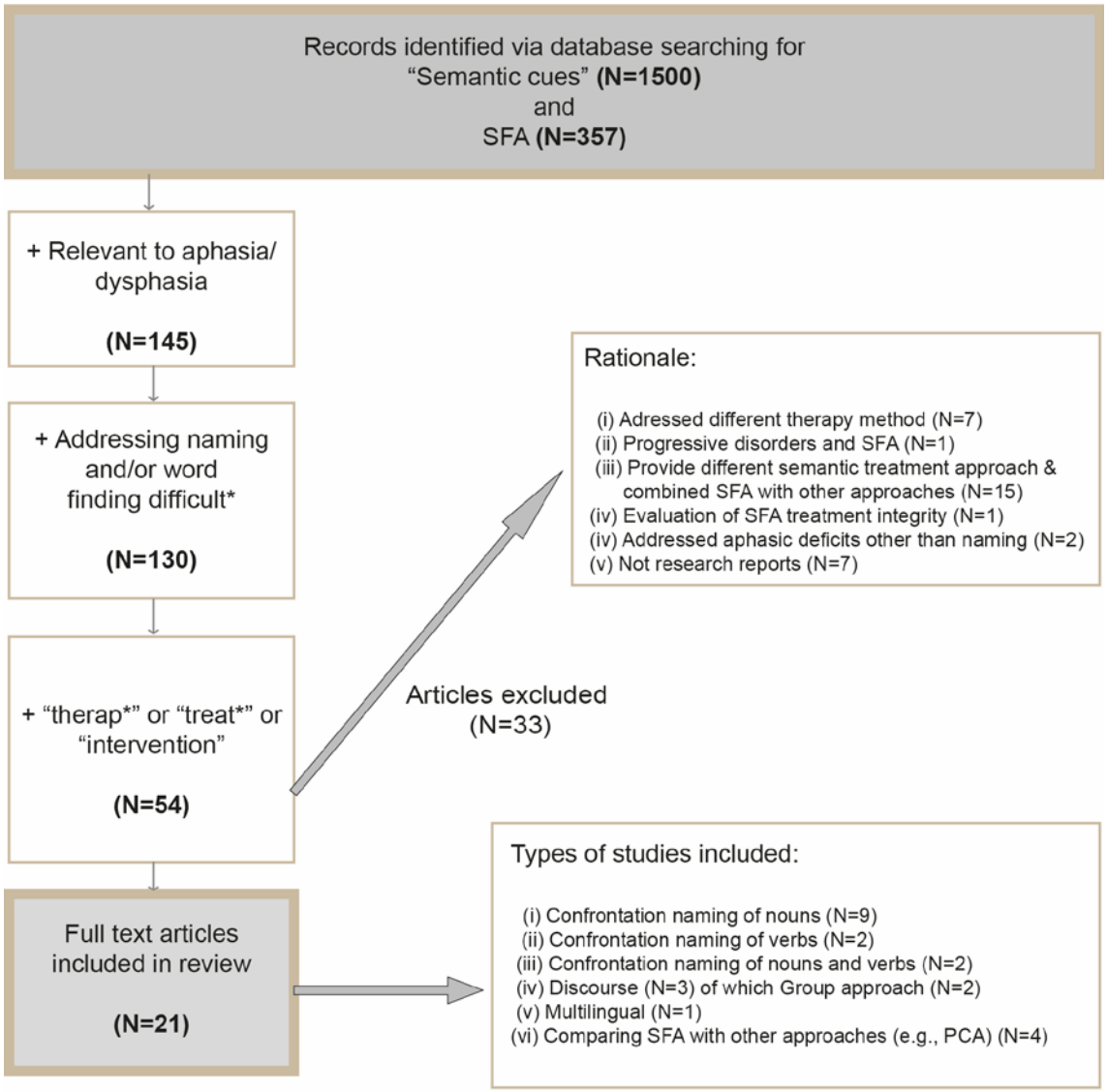


Boyle, 2004; Coelho et al., 2000



Wambaugh & Ferguson, 2007; Wambaugh et al., 2014

Figure 2: Identification process of articles from electronic databases



SFA: Semantic Features Analysis

PCA: Phonological Components Analysis

Table 1: Critical appraisal and methodological quality of studies (n=17) based on Single Case Experimental Design Scale (SCED)

Items of SCED Scale	Clinical History	Target Behaviours	Design	Baseline	Treatment Phase	Raw Data Record	Inter-Rater Reliability	Independence of Assessors	Statistical Analysis	Replication	Generalization	Total Score of SCED Scale	Phase of treatment
1. Massaro & Tomkins, 1994	YES	YES	MBAB	YES	YES	YES	YES	YES	NO	YES	YES	10	Pre-efficacy 1
2. Boyle & Coelho, 1995	YES	YES	ABA	YES	YES	YES	YES	NO	NO	NO	YES	8	Pre-efficacy 1
3. Coelho et al., 2000	YES	YES	ABA	YES	YES	YES	YES	NO	NO	NO	YES	8	Pre-efficacy 1
4. Boyle, 2004	YES	YES	MBAB	YES	YES	YES	YES	YES	YES	YES	YES	11	Pre-efficacy 1
5. Davis & Stanton, 2005	YES	YES	MBAB	YES	YES	YES	YES	NO	NO	NO	YES	8	Pre-efficacy 1
6. Wambaugh & Ferguson, 2007	YES	YES	MBAB	YES	YES	YES	YES	YES	YES	YES	NO	10	Pre-efficacy 1
7. Rider et al., 2008	YES	YES	MBAB	YES	YES	YES	YES	YES	YES	YES	NO	10	Efficacy
8. Antonucci, 2009	YES	YES	ABA	YES	YES	YES	YES	YES	YES	YES	Partly	10.5	Pre-efficacy 2
9. Marcotte & Ansaldo, 2010	YES		AB		Not a single case study but an observation control study						No		Pre-efficacy 1
10. Peach & Reuter, 2010	YES	YES	Single case time series across behaviors	YES	YES	YES	YES	YES	YES	YES	Variable	10.5	Pre-efficacy 1
11. Hashimoto & Frome, 2011	YES	YES	MBAB	YES	YES	YES	YES	YES	YES	YES	YES	11	Pre-efficacy 1
12. Falconer & Antonucci, 2012	YES	YES	ABA	YES	YES	YES	YES	NO	YES	YES	YES	9	Pre-efficacy 2

Items of SCED Scale	Clinical History	Target Behaviours	Design	Baseline	Treatment Phase	Raw Data Record	Inter-Rater Reliability	Independence of Assessors	Statistical Analysis	Replication	Generalization	Total Score of SCED Scale	Phase of treatment
13. Hashimoto, 2012	**YES**	**YES**	**MBAB**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**Partly**	**10.5**	Pre-efficacy 2
14. van Hees et al., 2013	**YES**	**YES**	**ABA**	**YES**	**YES**	**YES**	**NO**	**NO**	**YES**	**YES**	**NO**	**8**	Pre-efficacy 2
15. Wambaugh et al., 2014	**YES**	**YES**	**MBAB**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**NO**	**10**	Pre-efficacy 2
16. Kristensson et al., 2015	**YES**	**YES**	**MBAB**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**NO**	**10**	Pre-efficacy 1
17. DeLong et al., 2015	**YES**	**YES**	**MBAB**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**Variable**	**10.5**	Pre-efficacy 1
18. Knoph et al.,2015	**YES**	**YES**	**AB**	**YES**	**YES**	**YES**	**YES**	**NO**	**YES**	**YES**	**YES**	**10**	Pre-efficacy 2
19. Mehta & Isaki, 2016	**YES**	**YES**	**ABA**	**YES**	**YES**	**YES**	**NO**	**YES**	**YES**	**YES**	**NO**	**9**	Pre-efficacy 2
20. Neumann, 2017	**YES**	**YES**	**MBAB**	**YES**	**YES**	**YES**	**NO**	**NO**	**YES**	**YES**	**YES**	**9**	Pre-efficacy 2
21. Sadeghi et al., 2017	**YES**	**YES**	**AB**	**YES**	**YES**	**YES**	**NO**	**NO**	**YES**	**YES**	**YES**	**8**	Pre-efficacy 2

SCED: Single Case Experimental Design
MBAB: Multiple baseline across behaviors study, involving multiple assessments pre- treatment, post-treatment and follow up
AB: Pre- / post- treatment study
ABA: Pre- / post- treatment / follow up study

Table 2: Study characteristics: number of participants, type of SFA treatment, dosage, duration and amount of treatment

Study	n	Type of SFA	Language	Treatment dosage and duration	Total amount of treatment (mins)
1. Massaro & Tompkins, 1994	2	Noun SFA	Monolingual	21 sessions	CNC
2. Boyle & Coelho, 1995	1	Noun SFA	Monolingual	3*60min sessions/wk 6 weeks	1080
3. Coelho et al., 2000	1	Noun SFA	Monolingual	3*60min sessions/wk 7 weeks	1260
4. Boyle, 2004	2	Noun SFA	Monolingual	3*50- 75 min sessions/wk 8 weeks	≈1500
5. Davis & Stanton, 2005	1	Noun SFA	Monolingual	2* 60 min sessions/wk 6 weeks	720
6. Wambaugh & Ferguson, 2007	1	Verb SFA	Monolingual	3*45 - 60 min sessions/wk 4 weeks	≈630
7. Rider et al., 2008	3	Noun SFA	Monolingual	2-3 * 60min sessions/wk 5 weeks or 80% naming accuracy across 2 sessions	≈750
8. Antonucci, 2009	3	Group Approach Discourse SFA	Monolingual	2*60 -90min sessions/wk 7 weeks	≈1050
9.Marcotte & Ansaldo, 2010	1	Nouns & Verb SFA	Monolingual	3*60min sessions/wk 3 weeks	540
10. Peach and Reuter, 2010	2	Discourse SFA	Bilingual	P1: 14 *50 min per sessions 10 weeks P2: 13*50 min per sessions 10 1⁄2 weeks	≈675
11. Hashimoto & Frome, 2011	1	Modified Noun SFA	Monolingual	2*60min sessions/wk 12 weeks	1440
12. Falconer & Antonucci, 2012	4	Group Approach Discourse SFA	Monolingual	2* 90 - 120 min sessions/wk 7 weeks & daily practice of homework	≈1470

Study	n	Type of SFA	Language	Treatment dosage and duration	Total amount of treatment (mins)
13. Hashimoto, 2012	2	SFA vs PCA	Monolingual	2*45-60min sessions per day 4 sessions/wk until >80% naming accuracy across 3 sessions 2 – 7 ½ weeks	≈1470
14. van Hees et al.,2013	8	SFA vs PCA	Monolingual	3* 45-90min sessions/wk 4 weeks	≈810
15. Wambaugh et al., 2014	4	Verb SFA	Monolingual	3*60min sessions/wk Until 90% accurate naming of trained items in 2-3 probes or 4 weeks	720
16. Kristensson et al., 2015	3	Nouns & Verb SFA	Monolingual	20* 60min sessions 5-6 weeks	≈1200
17. DeLong et al., 2015	5	Noun SFA	Monolingual	3* 50 min sessions/wk Max 20 treatment sessions per treatment phase or 86% items correct in 2 of 3 consecutive probe sessions	1000
18. Knoph et al.,2015	1	Verb Quadrilingual SFA	Quadrilingual	29 sessions 3 days per week 2.5 weeks	1320
19. Mehta & Isaki, 2016	2	Noun SFA	Monolingual	2*60min sessions/wk 8 weeks	720
20. Neumann, 2017	4	SFA vs PCA	2 Monolingual 2 Bilingual	2-3 * 120 min sessions/wk Max 10 trained items given or 40% or more above baseline in naming improvement in the treated items on 2 consecutive probe sessions 2 to 6 ½ weeks	CNC
21. Sadeghi et al., 2017	4	SFA vs PCA	Monolingual	7*45 min sessions 2 weeks	315

CNC: Cannot calculate

Table 3: Participants' demographic and stroke and aphasia characteristics (N=51)

Study	n	Participants	Age (years)	Gender	Etiology	TPO (months)	WAB AQ Aphasia Severity[a]	Aphasia Type	Fluency
1. **Massaro & Tompkins, 1994**	2	P1	24	M	TBI	60	NR	Broca	Non – Fluent
		P2	28	F	TBI	144	NR	NR	Non -Fluent
2. **Boyle &Coelho, 1995**	1	P1	57	M	L CVA	65	82 Mild	Broca	Non - Fluent
3. **Coelho et al., 2000**	1	P1	52	M	TBI	17	56.6 Moderate	NR	Fluent
4. **Boyle, 2004**	2	P1	70	M	L CVA	15	90.6 Mild	Anomic	Fluent
		P2	80	M	LCVA	14	61.2 Moderate	Wernicke	Fluent
5. **Davis & Stanton, 2005**	1	P1	59	F	CVA	4	102[b] Moderate	NR	Fluent
6. **Wambaugh & Ferguson, 2007**	1	P1	74	F	L CVA	50	67.7 Moderate	Anomic	Non - Fluent
7. **Rider et al., 2008**	3	P1	73	M	L CVA	26	74.6 Moderate - Mild	Transcortical Motor	Non – Fluent
		P2	55	F	L CVA	45	76.5 Mild	Transcortical Motor	Non – Fluent
		P3	62	M	L CVA	126	66 Moderate	Broca	Non - Fluent
8. **Antonucci, 2009**	3	P1	NR	M	NR	NR	NR	NR	NR
		P2	53	M	NR	18	63 Moderate	Conduction	Fluent
		P3	59	F	NR	16	90.2 Mild	NR	Fluent
9. **Marcotte & Ansaldo, 2010**	1	P1	66	M	CVA	84	Severe	Broca	Non – Fluent
10. **Peach & Reuter, 2010**	2	P1	77	F	L CVA	4	90.2 Mild	Anomic	Fluent
		P2	62	F	L CVA	14	70.3 Moderate	Anomic	Fluent
11. **Hashimoto & Frome, 2011**	1	P1	72	F	CVA	NR	35 Severe	Broca	Non -Fluent
12. **Falconer & Antonucci, 2012**	4	P1	35	M	M CVA	72	69.6 Moderate	Conduction	Fluent
		P2	55	M	L CVA	156	61 Moderate	Conduction	Fluent
		P3	31	M	TBI	96	34 Severe	Broca	Non –Fluent
		P4	62	F	M CVA	25	52.4 Moderate	Transcortical Motor	Non -Fluent
13. **Hashimoto, 2012**	2	P1	66	F	L CVA	60	49.5 Severe - Moderate	Wernicke	Non – Fluent
		P2	33	F	L CVA	18	57.5 Moderate	Broca	Fluent
14. **van Hees et al., 2013**	8	P1	60	F	L CVA	38	77.2 Mild – Moderate	Conduction	Fluent
		P2	60	M	L CVA	57	87.4 Mild	Anomic	Fluent
		P3	41	F	L CVA	170	92 Mild	Anomic	Fluent
		P4	52	F	L CVA	55	86.4 Mild	Conduction	Fluent
		P5	56	F	L CVA	25	57.3 Moderate	Anomic	Fluent
		P6	48	F	L CVA	17	81.7 Mild	Anomic	Fluent
		P7	69	M	L CVA	36	73.4 Moderate	Anomic	Fluent

Study	n	Participants	Age (years)	Gender	Etiology	TPO (months)	WAB AQ Aphasia Severity[a]	Aphasia Type	Fluency
		P8	65	M	L CVA	20	82.9 Mild	Anomic	Fluent
15. **Wambaugh et al., 2014**	4	P1	48	F	L MCA	276	77.4 Mild	Conduction	Fluent
		P2	53	M	L PCA	66	83.4 Mild	Anomic	Fluent
		P3	55	M	L CVA	79	53 Moderate	Broca	Non – Fluent
		P4	60	M	R MCA L MCA	21	66.9 Moderate	Broca	Non - Fluent
16. **Kristensson et al., 2015**	3	P1	71	M	L PCA	36	Moderate – Severe	Wernicke	Fluent
		P2	54	F	L BG	60	Moderate – Severe	Mixed	Non- Fluent
		P3	64	M	L MCA	24	Mild – Moderate	Broca	Non – Fluent
17. **DeLong et al., 2015**	5	P1	62	F	L CVA	11	64.5 Moderate	Conduction	Fluent
		P2	54	M	L MCA	30	58.3 Moderate	Wernicke	Fluent
		P3	30	M	L MCA	23	66 Moderate	Broca	Fluent
		P4	53	F	L MCA	384	78.4 Moderate	Anomic	Fluent
		P5	65	F	L MCA	12	18 Very Severe	Global	Non – Fluent
18. **Knoph et al., 2015**	1	P1	59	F	L NR	7	Moderate[c]	NR	Non – Fluent
19. **Mehta & Isaki, 2016**	2	P1	58	M	L CVA	108	53 Moderate	Wernicke	Fluent
		P2	58	M	L CVA	132	60.2 Moderate	Conduction	Fluent
20. Neumann, 2017	4	P1	41	M	L CVA	96	Moderate[d]	Conduction	Fluent
		P2	38	F	L CVA	24	Mild[d]	Anomic	Fluent
		P3	60	M	L NR	84	Mild[d]	Anomic	Fluent
		P4	47	M	L NR	24	Severe[d]	Anomic	Fluent
21. Sadeghi et al., 2017	4	P1	61	M	L CVA	24	NR	Broca	Non – Fluent
		P2	52	F	L CVA	17	NR	Broca	Non – Fluent
		P3	45	M	L MCA	67	NR	Anomic	Fluent
		P4	47	M	L CVA	15	NR	Broca	Non – Fluent

a: Aphasia severity based on Western Aphasia Battery-Revised (Kertesz, 2007) Aphasia Quotient. Retrieved October 1, 2015, from http://www.pearsonclinical.com/language/products/100000194/western-aphasia-batteryrevised.html

b: Based on Aphasia Diagnostic Profiles score (Helm-Estabrooks, 1992)

c: Based on Bilingual Aphasia Test (Paradis, Libben, & Hummel, 1987)

d: Based on Boston Diagnostic Aphasia Examination – Short Form (Goodglass, Kaplan &Barresi, 2001)

NR: not reported; R: Right hemisphere; L: left hemisphere; TPO: Time Post Onset; MCA: Middle Cerebral Artery; CVA: Cerebral Vascular Accident; PCA: Posterior Cerebral Artery; BG: Basal Ganglia

Table 4: Summary of treatment outcomes

Study	n	Treated items improved?	Maintenance	Generalization to untreated items?
1. Massaro & Tompkins, 1994	2	YES	YES	YES
		YES	YES	YES
2. Boyle & Coelho 1995	1	YES	YES	YES
3. Coelho et al., 2000	1	YES	YES	YES
4. Boyle, 2004	2	YES	YES	YES
		YES	Unavailable	YES
5. Davis & Stanton, 2005	1	YES	YES	YES
6. Wambaugh & Ferguson, 2007	1	YES	YES	NO
7. Rider et al., 2008	3	YES	YES	NO
		YES	YES	NO
		YES	NO	NO
8. Antonucci, 2009	3	YES	YES	YES
		YES	YES	NO
		NO	Unavailable	Unavailable
9. Marcotte & Ansaldo, 2010	1	YES		
10. Peach and Reuter, 2010	2	YES	NO	Variable
		YES	NO	Variable
11. Hashimoto & Frome, 2011	1	YES	YES	YES
12. Falconer & Antonucci, 2012	4	YES	YES	YES
		YES	YES	YES
		YES	YES	YES
		YES	YES	YES
13. Hashimoto, 2012	2	YES	YES	NO
		YES	YES	YES
14. van Hees et al., 2013	8	YES	YES	NO
		YES	NO	NO
		YES	YES	NO
		YES	YES	NO
		NO	NO	NO
		NO	NO	NO
		NO	NO	NO
		NO	NO	NO
15. Wambaugh et al., 2014	4	YES	YES	NO
		YES	YES	NO
		YES	YES	NO

Study	n	Treated items improved?	Maintenance	Generalization to untreated items?
		NO	NO	NO
16. Kristensson et al. 2015	3	NO	NO	NO
		NO	NO	NO
		NO	NO	NO
17. DeLong et al., 2015	5	YES	YES	NO
		YES	NO	NO
		YES	YES	NO
		YES	NO	NO
		NO	NO	NO
18. Knoph et al., 2015	1	YES		NO
19. Mehta & Isaki, 2016	2	YES	YES	
		YES	YES	
20. Neumann, 2017	4	YES	YES	YES
		YES	YES	YES
		YES	YES	YES
		YES	YES	YES
21. Sadeghi et al., 2017	4	YES		YES
		YES		YES
		YES		YES
		YES		YES
Total	55	YES n=45 (81.82%) NO n=10 (18.19%)	YES n=32 (58.18%) NO n=15 (27.27%) Unavailable n=2 (3.63 %) NR n=6	YES n=22 (40%) NO n=27 (49.09%) Variable n=2 (3.63%) Unavailable n=1 (1.81%)

Table 5: Time of Assessments after Therapy

Study	Number of Assessments	Time of Assessment after Therapy				
1. Massaro & Tompkins, 1994	2	Immediately after therapy	2 weeks			
2. Boyle & Coelho 1995	3	Immediately after therapy	1 month	2 months		
3. Coelho et al., 2000	3	Immediately after therapy	1 month	2 months		
4. Boyle, 2004	2	Immediately after therapy	1 month			
5. Davis & Stanton, 2005	5	Immediately after therapy	6 weeks	12 weeks	18 weeks	1 year
6. Wambaugh & Ferguson, 2007	3	Immediately after therapy	2 weeks	6 weeks		
7. Rider et al., 2008	2	Immediately after therapy	4 weeks			
8. Antonucci, 2009	2	Immediately after therapy	6 weeks			
9. Marcotte & Ansaldo, 2010	1	Immediately after therapy				
10. Peach and Reuter, 2010	2	Immediately after therapy	4 ½ months			
11. Hashimoto & Frome, 2011	2	Immediately after therapy	6 weeks			
12. Falconer & Antonucci, 2012	2	Immediately after therapy	6 weeks			
13. Hashimoto, 2012	2	Immediately after therapy	6 weeks			
14. van Hees et al., 2013	2	Immediately after therapy	2-3 weeks			
15. Wambaugh et al., 2014	3	Immediately after therapy	2 weeks	6 weeks		
16. Kristensson et al. 2015	2	Immediately after therapy	10-12 weeks			
17. DeLong et al., 2015	3	Immediately after therapy	2 weeks	6 weeks		
18. Knoph et al., 2015	1	Immediately after therapy				
19. Mehta & Isaki, 2016	2	Immediately after therapy	8 weeks			
20. Neumann, 2017	2	Immediately after therapy	4-6 weeks			
21. Sadeghi et al., 2017	1	Immediately after therapy				

Table 6: Clinical Efficacy: effect sizes and percent of non-overlapping data.

Study	Participants	Cohen's d	PND	Magnitude of effect
1. Massaro & Tompkins, 1994	P1	7.45[c]		Medium effect
	P2	3.54[c]		Less than small effect
2. Boyle & Coelho 1995	P1	16.36[c]		Large effect
3. Coelho et al., 2000	P1	4.41[c]		Small effect
4. Boyle, 2004	P1	18.48[a]		Large effect
	P2	CNC	100%	Highly effective
5. Davis & Stanton, 2005	P1	CNC	91.67%	Highly effective
36. Wambaugh & Ferguson, 2007	P1	6.35[a]		Small effect
7. Rider et al., 2008	P1	3.86[b]		Less than small effect
	P2	5.54[b]		Small effect
	P3	2.97[b]		Less than small effect
8. Antonucci, 2009	P1	CNC	CNC	
	P2	ns	CNC	
	P3	2.05[b]	CNC	Less than small effect
9. Marcotte & Ansaldo, 2010	P1	CNC	CNC	-
				-
10. Peach & Reuter, 2010	P1	1.79		Less than small effect
	P2		85%	Moderate effective
11. Hashimoto & Frome, 2011	P1	10.56[b]		Large effect
12. Falconer & Antonucci, 2012	P1	3.44		Less than small effect
	P2	4.16		Small effect
	P3	0.03		Less than small effect
	P4	1.28		Less than small effect
13. Hashimoto, 2012	P1	7.11		Medium effect
	P2	7		Medium effect
14. van Hees et al., 2013	P1	5.86		Small effect
	P2	3.79		Less than small effect
	P3	4.54		Small effect
	P4	7.79		Medium effect
	P5	ns		-
	P6	ns		-
	P7	ns		-
	P8	ns		-
15. Wambaugh et al., 2014	P1	6.87[b]		Small effect
	P2	13.14[b]		Large effect
	P3	1.58[b]		Less than small effect
	P4	8.53[b]		Medium effect
16. Kristensson et al., 2015	P1	1.06[b]		Less than small effect
	P2	0.66[b]		Less than small effect
	P3	0.64[b]		Less than small effect
17. DeLong et al., 2015	P1	3.03[b]		Less than small effect
	P2	2.20[b]		Less than small effect
	P3	4.68[b]		Small effect
	P4	6.66[b]		Small effect
	P5	CNC	CNC	-

18. Knoph et al., 2015	P1	10.07		Large effect
19. Mehta & Isaki, 2016	P1	CNC	CNC	-
	P2	CNC	CNC	-
20. Neumann, 2017	P1	1.12[b]		Less than small effect
	P2	9.58[b]		Medium effect
	P3	1.67[b]		Less than small effect
	P4	11.21[b]		Large effect
21. Sadeghi et al., 2017	P1	5.08 [b]		Small effect
	P2	6.89 [b]		Small effect
	P3	19.23[b]		Large effect
	P4	11.89[b]		Large effect

PND: percent of non-overlapping data; CNC: Cannot calculate, [a]: Calculated by first author of this paper, [b]: Average calculation by first author of this paper, [c] : Calculated by first author of this paper based on substituted data from other phases, [ns]: no substantial change