



Available online at ScienceDirect www.sciencedirect.com





Annals of Physical and Rehabilitation Medicine 57 (2014) 254-267

Literature review / Revue de la littérature

The effectiveness of semantic feature analysis: An evidence-based systematic review

L'efficacité de l'analyse des traits sémantiques : une revue systématique fondée sur des preuves

K.M. Maddy ^{a,*}, G.J. Capilouto ^b, K.L. McComas ^c

Abstract

Objectives. – This review examines the effectiveness of semantic feature analysis as an intervention to improve naming abilities for persons with aphasia.

Method. – A systematic search of the literature identified 11 studies that met the pre-determined inclusion criteria. Two independent raters evaluated each study for methodological quality and assigned appropriate levels of evidence using the Single Case Experimental Design scale. To determine clinical effectiveness, effect sizes using Cohen's d were calculated if sufficient data were available. Alternatively, percent of non-overlapping data was calculated.

Results. – Results indicated that methodologically sound research has been conducted to determine the effectiveness of semantic feature analysis for persons with aphasia using single subject research designs. When using Cohen's d, the majority of participants showed a small effect size. However, when percent of non-overlapping data was calculated, a large treatment effect was present for the majority of participants.

Conclusions. – Semantic feature analysis was an effective intervention for improving confrontational naming for the majority of participants included in the current review. Further research is warranted to examine generalization effects.

© 2014 Elsevier Masson SAS. All rights reserved.

Keywords: Semantic feature analysis; Aphasia; Anomia; Treatment; Systematic review

Résumé

Objectifs. – Cette revue examine l'efficacité de l'analyse des traits sémantiques comme intervention visant à améliorer les capacités de désignation de personnes atteintes d'aphasie.

Méthodes. – Une recherche systématique de la littérature a repéré 11 études correspondant à des critères d'inclusion prédéterminés. Deux évaluateurs indépendants ont noté chaque étude en termes de qualité méthodologique et de niveau de preuve en utilisant l'échelle Single Case Experimental Design (SCED). Afin de déterminer le degré d'efficacité clinique, des ampleurs de l'effet utilisant le d de Cohen étaient calculées à condition de disposer d'un nombre suffisant de données disponibles. Sinon, le pourcentage de données mutuellement exclusives était calculé. Résultats. – Les résultats ont indiqué que des recherches méthodologiquement valables avaient été conduites en vue de déterminer l'efficacité de l'analyse des traits sémantiques chez les aphasiques, analyse qui appliquait des plans de recherche centrés sur un sujet unique. Lors de l'utilisation du d de Cohen, la majorité des participants n'ont présenté qu'une petite ampleur de l'effet. Or dès que le pourcentage de données mutuellement exclusives a été calculé, un effet traitement important a été constaté chez la plupart des participants.

E-mail address: klmc227@uky.edu, katiemmaddy@gmail.com (K.M. Maddy).

^a Department of Rehabilitation Sciences, University of Kentucky, Charles T. Wethington Building 124A, 900 S. Limestone, Lexington, KY 40536-0200, United States

^b Department of Rehabilitation Sciences, University of Kentucky, Charles T. Wethington Building 124G, 900 S. Limestone, Lexington, KY 40536-0200, United States

^c Department of Communication Disorders, Marshall University, 109 Old Main, 1 John Marshall Drive, Huntington, WV 25755, United States Received 27 August 2013; accepted 11 March 2014

^{*} Corresponding author.

Conclusions. – L'analyse des traits sémantiques a constitué une intervention efficace dans le cadre de l'amélioration de la désignation par rapport au mot cible (confrontational naming) de la majorité des participants figurant dans cette revue. D'autres recherches pourraient examiner des effets de généralisation.

© 2014 Elsevier Masson SAS. Tous droits réservés.

Mots clés : Analyse des traits sémantiques ; Aphasie ; Anomie ; Traitement ; Recherche systématique

1. English version

1.1. Introduction

Anomia is the most frequent and persisting symptom of aphasia [1,2]. Anomia occurs secondary to a neurological event (e.g., stroke, brain injury) and inhibits the ability to formulate language, even at the most basic word level. Anomia is a defining feature of aphasia; it extends across multiple subtypes of aphasia and is observed for all grammatical word forms (e.g., nouns, verbs) [3,4]. Word naming deficits negatively affect people's ability to communicate their wants and needs and engage in important social settings and activities. In fact, prior research indicates that people with aphasia (PWA) tend to be more troubled by impairments in speaking than in reading, writing, or listening comprehension and impairments in speaking have important effects on how PWA are regarded by others in daily life [2,4]. Therefore, identifying an effective treatment for the improvement of naming deficits in PWA is critical.

There are a variety of treatment approaches focused on improving expressive language abilities, specifically anomia, in PWA. It is believed that anomia typically results from overarching semantic impairments and reflects an insufficient engagement of the critical features that distinguish concepts from one another [5]. Typically, semantic approaches to treatment are used to target anomia as opposed to phonological approaches. Examples include circumlocution-induced naming [6], personalized cueing [7–9], and semantic feature analysis (SFA) [10–13]. The focus of the current review is the clinical effectiveness of SFA for the treatment of anomia in adults with neurological injury.

SFA was first introduced by Ylvisaker and Szekeres [14] and later refined by Massaro and Tompkins [15]. It is a commonly used treatment to improve naming and expressive language abilities of PWA by providing an organized method of activating semantic networks [16]. SFA uses a systematic cueing technique whereby PWA are asked to produce words semantically related to the target word they cannot recall [5]. For example, if the target word is "cup", the cues might involve questions related to its use (e.g., What do you do with it?), its properties (e.g., What does it look like?), where it might be used (e.g., Where do you find it?), what category it belongs to, and what might be associated with it (e.g., What are other things that are similar to it?). Because it is suggested that anomia results from an impaired semantic network, the goal of therapy is to alter the semantic network connectivity through refinement of the damaged network [5]. Hypothetically, SFA improves the retrieval of conceptual information by accessing and refining semantic networks [15]. Increased activation of the semantic

network surrounding the target word elevates the likelihood the word will be retrieved and may also aid to repair the damaged semantic system [5,12,15].

A recent review by Boyle [17] examined the effectiveness of SFA. The review included seven studies, and all studies reported positive outcomes for the effectiveness of SFA for improving anomia for individuals with neurological impairments. However, only three reported the magnitude of the treatment effect [18–20]. A significant limitation of Boyle's [17] review is that there was no attempt to calculate magnitude of effect using data reported in the included studies. The absence of effect size calculations in the remaining studies made it difficult to conclude the effectiveness of the treatment. Therefore, the aim of the current study is two-fold. First, we extend Boyle's [17] review by including new research. Second, we apply statistical methods to investigate the magnitude of treatment effect for the included studies in an effort to answer the following clinical question: For patients with nondegenerative aphasia, does semantic feature analysis improve confrontational naming abilities?

1.2. Methods

A search of the literature was conducted to identify studies that investigated SFA as the primary intervention for anomia for PWA. Seven electronic databases were searched through June 2013: Academic Search Premier, AgeLine, CINAHL, ERIC, Medline, PsycInfo, and Linguistics and Language Behavior Abstracts. An additional search was performed within the American Speech-Language and Hearing Association (ASHA) journals, and references from all relevant articles were examined to identify any other applicable studies. A combination of search terms included: *aphasia*, *semantic feature analysis*, *language disorder*, *semantic cues*, *anomia*, *language treatment*, and *naming treatment*.

Eighty-eight articles were identified through the search process and the first author subjected these to a title and abstract review, which eliminated 70 articles. A study was excluded from the review if it was not experimental in nature, did not address SFA as a treatment and did not include adults with neurological damage. Studies were considered for review if they were written in English and published in a peer-reviewed journal between 1980 and June 2013. The first author reviewed the remaining 18 articles and an additional 7 were excluded as they combined SFA with other treatment methods (e.g., group therapy or semantic priming therapy) or used a hybrid approach to SFA. To be true to the objective of this review, only those studies that used SFA as initially designed by Massaro and Tompkins [15] were included. The process for identifying

articles to be included in the evidence-based systematic review is displayed in Fig. 1.

Two certified and licensed speech-language pathologists critically evaluated the remaining studies for methodological quality. Ten studies were evaluated using the Single Case Experimental Design (SCED) scale [21]. The Marcotte and Ansaldo [22] study was not evaluated using the SCED as it was not a single subject research design study. The SCED is an 11point scale used to evaluate the methodological quality of single case experimental studies. A perfectly designed and executed study would receive a score of 11. Each independent rater provided a score of 1 if the study adequately addressed the specified quality item or a 0 if it was poorly addressed or not addressed at all. Following independent scoring, inter-rater reliability was calculated using Cohen's weighted kappa statistic [23]. The weighted kappa is the proportion of agreement beyond chance and takes into consideration the degree of disagreement between two independent raters. The weighted kappa score was .656, which is indicative of good inter-rater reliability [24]. When disagreements were present, an average score was calculated. The first author randomly selected three studies, and re-calculated SCED scores to determine intra-rater reliability. To reduce bias and ensure ratings were not dependent upon one another, re-scoring was completed two weeks after the initial scoring [25]. The weighted kappa score for intra-rater reliability was 1.0, indicating perfect intra-rater reliability.

It is well understood that clinical effectiveness is a critical variable to consider for any treatment. At the same time, it is important to interpret clinical effectiveness in the context of a given study. To strengthen the interpretation of the result of this systematic review, levels of evidence, based on the American Speech and Hearing Association's (ASHA) levels of evidence hierarchy, were also assigned. The ASHA hierarchy was chosen due to its applicability to the effectiveness of interventions for individuals with communication disorders. Two independent raters evaluated each study and assigned a level of evidence. Following independent rating, the weighted kappa was calculated to assess reliability for assigning levels. The first author randomly chose 30% of the studies for reevaluation, to calculate intra-rater reliability. Inter- and intra-rater reliability were excellent (k = 1.0).

To determine the clinical efficacy of SFA as a treatment method for aphasia, effect sizes were calculated in those studies that reported sufficient data. Prior to calculation, it was necessary to determine the individual values for the pretreatment and post-treatment phases for each set of trained items. Therefore, effect size was calculated when response data were presented in whole number increments using a variation of Cohen's *d* statistic [27] as described by Busk and Serlin [28]. The following formula was used:

$$d_1 = \frac{xA_2 - xA_1}{sA_1}$$

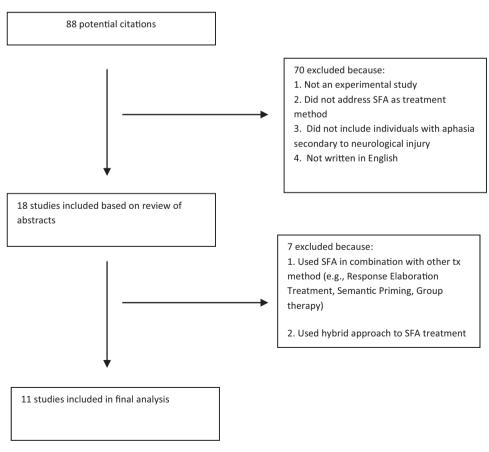


Fig. 1. Process for identification of included studies.

where A_2 and xA_1 designate post-treatment and pre-treatment time periods, respectively, xA is the mean of the data collected during the specific period, and sA_1 is the corresponding standard deviation. Although three post-treatment probes are recommended to provide a more reliable estimate of magnitude of effect, it is mathematically possible to calculate effect sizes when only one post-treatment probe is reported [29]. For the purpose of the present study, effect sizes were calculated when at least one post-treatment probe was reported. The magnitude of change in level of performance was determined according to the benchmarks for lexical retrieval studies described by Beeson and Robey [29]. The benchmarks were 4.0, 7.0, and 10.1 for small, medium, and large effect sizes, respectively.

When it was not possible to calculate the d statistic, 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 [30,31]. PND is the percentage of phase B data points (the treatment phase) that do not overlap (in a favorable direction) with phase A data points (baseline or no treatment). To determine the magnitude of effect, benchmarks put forth by Scruggs et al. [32] were used. A PND greater than 90% was interpreted to reflect a highly effective treatment; a PND of 70–90% was considered a moderate treatment

outcome; and, a PND of 50–70% was considered a questionable effect. Any PND of less than 50% is interpreted as an ineffective intervention since performance during intervention has not affected behavior beyond baseline performance.

1.3. Results

Ten studies included in the review were single subject research design studies and one was a pre- and post-treatment study. Twenty-four participants were represented in the review. Demographic variables of interest for all participants are presented in Table 1. There was significant heterogeneity in age and time post onset of participants. Participants ranged in age from 24 to 85 years, with a mean age of 60.75 (S.D. = 14.29). Time post onset ranged from 4 to 187 months, with a mean of 53.74 months (S.D. = 50.44). There were 15 males (63%) and 9 females (37%) represented. Non-fluent aphasia was the most represented subtype in this review (n = 17; 71%) and seven participants (29%) had fluent aphasia. Twenty-one participants had sustained a single cerebrovascular accident. The remaining three participants presented with aphasia following traumatic brain injury. Of the 11 studies included in the review, aphasia severity was reported in only five studies.

Table 1 Participant demographic information.

Study	n	Participant	Age (years)	Gender	Etiology	TPO (months)	Aphasia type	Aphasia severity
Boyle and Coelho, 1995 [12]	1	P1	57	M	L CVA	65	Non-fluent	NR
Boyle, 2004 [11]	2	P1	70	M	L CVA	15	Fluent	NR
		P2	80	M	L CVA	14	Fluent	NR
Coelho et al., 2000 [13]	1	P1	52	M	TBI	17	Fluent	Moderate- severe
Davis and Stanton, 2005 [34]	1	P1	59	F	CVA	4	Fluent	NR
Hashimoto, 2011 [33]	1	P1	72	F	CVA	NR	Non-fluent	Moderate- severe
Marcotte and Ansaldo, 2010 [22]	1	P1	66	M	CVA	84	Non-fluent	Severe
Massaro and Tompkins, 1994 [15]	2	P1	24	M	TBI	60	Non-fluent	NR
		P2	28	F	TBI	144	Non-fluent	NR
Peach and Reuter, 2010 [19]	2	P1	77	F	L CVA	4	Fluent	Mild
		P2	62	F	L CVA	14	Fluent	Moderate
Rider et al., 2008 [18]	3	P1	73	M	L CVA	26	Non-fluent	NR
		P2	55	F	L CVA	45	Non-fluent	NR
		P3	62	M	L CVA	126	Non-fluent	NR
Wambaugh and Ferguson, 2007 [3]	1	P1	74	F	L CVA	50	Non-fluent	Moderate
Wambaugh et al., 2013 [20]	9	P1	85	M	CVA	126	Non-fluent	NR
		P2	59	F	CVA	42	Non-fluent	NR
		P3	61	M	CVA	31	Non-fluent	NR
		P4	47	M	CVA	187	Non-fluent	NR
		P5	59	M	CVA	65	Non-fluent	NR
		P6	52	M	CVA	13	Non-fluent	NR
		P7	66	M	CVA	65	Non-fluent	NR
		P8	64	F	CVA	18	Non-fluent	NR
		P9	54	M	CVA	9	Fluent	NR

NR: not reported; CVA: cerebrovascular accident; L: left hemisphere; TPO: time post onset.

Table 2 Critical appraisal of the literature.

Study	Research design	SCED quality score	
Boyle and Coelho, 1995 [12]	A-B single subject	8	
Boyle, 2004 [11]	Multiple baseline across behaviors	10	
Coelho et al., 2000 [13]	A-B single subject	8.5	
Davis and Stanton, 2005 [34]	Multiple baseline across behaviors	8	
Hashimoto, 2011 [33]	Multiple baseline across behaviors	8	
Marcotte and Ansaldo, 2010 [22]	Pre-/post-treatment study	N/A	
Massaro and Tompkins, 1994 [15]	Multiple baseline across behaviors	10	
Peach and Reuter, 2010 [19]	Single case time series across behaviors	10	
Rider et al., 2008 [18]	Multiple probes across behaviors	10.5	
Wambaugh and Ferguson, 2007 [3]	Multiple baseline across behaviors	10	
Wambaugh et al., 2013 [20]	Multiple baseline across behaviors	10	

Scores on the Single Case Experimental Design (SCED) ranged from 8.0 to 10.5 with an average score of 9.3 out of 11 (Table 2). Following SCED scoring, studies were assigned levels of evidence [26]. Ten studies were considered well-designed, quasi-experimental studies and received a level IIb rating. The Marcotte and Ansaldo [22] study was determined to be an observational controlled study and assigned a level III rating. The prevalence of high SCED scores and IIb evidence level ratings indicate strong methodological quality and rigor for the reviewed studies.

Three studies self-reported effect sizes for participants [3,18,19]. Effect sizes were calculated for five participants (Table 3). When data was collected and reported on two or more trials of the dependent variable, an average effect size was calculated [18,20,33]. A large effect size was present for two participants (d = 18.76; d = 10.56). A medium effect was present for six participants (d = 7.95; d = 9.28; d = 7.48; d = 7.79; d = 7.32; d = 9.617). A small effect size (d = 1.79; d = 3.86; d = 5.54; d = 2.97; d = 1.26; d = 5.01; d = 5.32; d = 6.35) was present for eight participants. The high

Table 3
Effect sizes and percent of non-overlapping data for included studies.

			PND	Magnitude of effect
P1	Non-fluent		100%	Highly effective
P1 P2	Non-fluent Fluent	18.48 ^b CND	100%	Large effect Highly effective
P1	Fluent	CND	100%	Highly effective
P1	Fluent	CND	91.67%	Highly effective
P1	Non-fluent	10.56 ^a		Large effect
P1	Non-fluent	CND	CND	N/A
P1 P2	Non-fluent Non-fluent	CND CND	100% 100%	Highly effective Highly effective
P1 P2	Fluent Fluent	1.79 CND 85%		Small effect Moderately effective
P1 P2 P3	Non-fluent Non-fluent Non-fluent	3.86 ^a 5.54 ^a 2.97 ^a		Small effect Small effect Small effect
P1	Non-fluent	6.35 ^b		Medium effect
P1 P2 P3 P4 P5 P6 P7 P8	Non-fluent Non-fluent Non-fluent Non-fluent Non-fluent Non-fluent Non-fluent Non-fluent	5.32 ^a 7.95 ^a 9.28 ^a 7.48 1.26 ^a 5.01 ^a 7.79 ^a 7.32 ^a		Small effect Medium effect Medium effect Medium effect Small effect Small effect Medium effect Medium effect
	P1 P2 P1 P1 P1 P1 P1 P2 P1 P2 P1 P2 P1 P2 P3 P1 P2 P3 P4 P5 P6 P7	P1 Non-fluent P2 Fluent P1 Fluent P1 Fluent P1 Non-fluent P1 Non-fluent P1 Non-fluent P2 Non-fluent P1 Fluent P2 Fluent P1 Non-fluent P2 Non-fluent P3 Non-fluent P1 Non-fluent P2 Non-fluent P3 Non-fluent P4 Non-fluent P5 Non-fluent P6 Non-fluent P7 Non-fluent P8 Non-fluent	P1 Non-fluent 18.48b P2 Fluent CND P1 Fluent CND P1 Fluent CND P1 Non-fluent 10.56a P1 Non-fluent CND P1 Non-fluent CND P1 Non-fluent CND P1 Fluent CND P1 Fluent CND P1 Non-fluent 3.86a P2 Fluent CND P1 Non-fluent 5.54a P3 Non-fluent 5.54a P3 Non-fluent 2.97a P1 Non-fluent 5.32a P2 Non-fluent 5.32a P3 Non-fluent 7.95a P3 Non-fluent 7.48 P5 Non-fluent 7.48 P5 Non-fluent 5.01a P6 Non-fluent 7.79a P8 Non-fluent 7.32a	P1 Non-fluent 18.48b P2 Fluent CND 100% P1 Fluent CND 100% P1 Fluent CND 91.67% P1 Non-fluent CND 91.67% P1 Non-fluent CND CND P1 Non-fluent CND 100% P2 Non-fluent CND 100% P1 Fluent CND 85% P1 Non-fluent 3.86a P2 Non-fluent 5.54a P3 Non-fluent 2.97a P1 Non-fluent 5.32a P2 Non-fluent 7.95a P3 Non-fluent 7.95a P4 Non-fluent 7.48 P5 Non-fluent 7.48 P6 Non-fluent 5.01a P7 Non-fluent 7.79a P8 Non-fluent 7.32a

PND: percent of non-overlapping data.

^a Average calculation.

^b Calculated by the author of the present paper, from data provided in articles.

prevalence of small effect sizes indicates that the treatment was not effective in improving anomia for the majority of participants as it was for these eight participants. Effect size and PND could not be calculated for one participant from the Marcotte and Ansaldo [22] study.

PND was calculated for seven participants. A large treatment effect (PND > 90%) was present for six participants and a moderate treatment effect was present for one participant (PND = 85%). When examining clinical effectiveness using PND, treatment was highly effective for the majority of participants. None of the participants evidenced PND scores consistent with ineffective treatment.

1.4. Discussion

Results indicate that SFA is an effective intervention for improving confrontational naming of items trained in therapy for individuals with non-degenerative aphasia. In addition to self-reported positive outcomes, effect size calculations indicated that there was a medium to large treatment effect for the majority of participants represented in the review. Although findings suggest that treatment was effective for improving confrontational naming of trained items, limited generalization to untrained items and connected speech were reported in many of the included studies. This finding warrants further examination to determine factors that may influence generalization, such as intensity or dosage. There are several limitations to this review, which will be discussed in detail.

1.4.1. Variations in study participants

For the purpose of this study, participants in the acute phase of recovery were included. Although spontaneous recovery could result in favorable treatment results, all participants in the acute phases demonstrated a stable pre-treatment baseline. Additionally, participants in the acute stages of recovery did not demonstrate an abnormally large effect size or PND following treatment, d = 1.79 [19], PND = 85% [19], and PND = 91.67% [34]. Although there was considerable variation in the reported time post onset amongst participants, 92% (22/24) were postacute (more than 6 months post onset). This is not surprising as most single subject research studies in the aphasia literature are conducted with individuals who are significantly post onset in order to achieve a stable series of observations and to control for spontaneous recovery [29]. As a result, the magnitude of effects reported likely underestimates the magnitude of effects possible for individuals with acute aphasia, as more progress and improvement is typically observed in the acute stage of recovery [5]. Furthermore, throughout the aphasia literature, age and time post onset are often cited as significant prognostic indicators for recovery of language abilities [5,35]. Future studies need to investigate the effectiveness of SFA for various age groups and time post onset of aphasia.

Aphasia type was varied (e.g., anomic, transcortical motor, Broca's) amongst participants in the study. When divided into the broad categories of fluent and non-fluent aphasia, non-fluent aphasia was the most represented subtype in this review (n = 17; 71%). Seven participants (29%) had fluent aphasia.

Results suggest that the effectiveness of SFA as treatment for anomia may be more effective for persons with fluent aphasias as compared to those with non-fluent aphasias. Previous research suggests that individuals with fluent aphasias demonstrate significant deficits in category knowledge [36,37]. Thus, treatments targeting category knowledge deficits may be more effective for this group. However, further research is necessary to compare the benefits of SFA for individuals with both fluent and non-fluent aphasias. Consequently, clinicians contemplating SFA as a treatment may need to consider aphasia type.

1.4.2. Variations in the intervention

Clinicians and researchers should bear in mind that SFA is not always conducted using a standardized procedure. For example, some studies reported longer treatment periods while some included more frequent treatment sessions over a shorter duration. In the Rider et al. [18] study, treatment consisted of one-hour sessions two to three times a week for five weeks. Davis and Stanton [34] used a similar treatment protocol with therapy two times a week for six weeks. Further research is warranted to determine the appropriate dosage and treatment duration to determine clinical efficacy. Additionally, some studies targeted atypical exemplars (e.g., egret), while others targeted more typical exemplars (e.g., robin). Standard execution of SFA following the original intent of the treatment, including stimuli and treatment dosage, would allow for the conduction of meta-analysis to more conclusively determine treatment effectiveness.

1.4.3. Variations in outcome measures

There was significant heterogeneity regarding outcome measures used in the included studies. Peach and Reuter [19] examined the effect of SFA on discourse production, specifically verbal productivity (# of words per T-unit) and informativeness (content information units; %CIUs) [38], as well as naming of untrained items. Because these measures were not directly related to the outcomes of the treatment, it is difficult to calculate meaningful effect sizes. Boyle and Coelho [12], Davis and Stanton [34], and Boyle [11] analyzed the percentage of pictures correctly named as the primary dependent variable in the study, but also calculated %CIUs. In addition to these measures, Boyle and Coelho [12] also used the Communicative Effectiveness Index [39] to examine social validity of changes. In addition to percentage of pictures named correctly, Rider et al. [18] also examined lexical diversity (D) and performance on standardized language measures, such as the Western Aphasia Battery [40], Boston Naming Test [41], and the Reading Comprehension Battery for Aphasia [42].

1.4.4. Limitations to statistical analysis

The use of a variety of outcome measures makes it difficult to make a definitive decision regarding the clinical impact of the treatment. Therefore, meta-analysis is needed. There are several advantages to completing a meta-analysis [43]. First, meta-analysis allows for the generation of a summary effect by combining studies and establishes a magnitude of treatment

effect. It also provides for a determination if the effect is uniform across studies, and if not, provides a mechanism for studying participants' heterogeneity. Lastly, meta-analysis also allows for comparisons across subgroups (e.g., gender, age, aphasia type, aphasia severity).

A recent review of single subject research design studies found that PND is the most commonly used outcome metric for systematic reviews [30]. However, there are certain limitations of using PND. It is important to note that there are no benchmarks for PND published for studies in aphasia, therefore, it is difficult to estimate magnitude of effect for aphasia treatment, and findings should be interpreted cautiously. Furthermore, we found that PND may overestimate the effect of treatment, when compared to effect size d calculations. For example, Rider et al. [18] reported small effect sizes for three participants. However, after calculating PND, we found a highly effective (PND = 90%), moderately effective (PND = 80%), and questionably effective score (PND = 50%). Similarly, Peach and Reuter [19] reported a small effect size (d = 1.79) for one participant on verbal productivity during the treatment phase. When PND was calculated, we found a moderate treatment effect (PND = 78%). These comparisons suggest that PND is a more liberal method of analyzing treatment effectiveness. It is important to remember that PND is the percentage of phase B data points (the treatment phase) that do not overlap (in a favorable direction) with phase A (baseline or no treatment) data points. An individual may improve naming ability by 10% on confrontational naming tasks while another individual may improve naming ability by 50%, but would receive the same PND scores if there was no overlap of treatment and baseline data points. Therefore, by its nature, this calculation determines whether there is a treatment effect, but does not fully examine the *magnitude* of effect. Future research is needed to determine the correlation between PND scores and the d statistic and to determine the reliability of PND as a measure of treatment magnitude in the field of aphasia treatments.

In order to complete a true meta-analysis, effect sizes are required. Because effect sizes are quantified in standard deviation units, they can be compared across studies and combined in meta-analyses [29]. Although effect sizes are now more commonly reported in group studies, they continue to be generally absent in single subject research design studies. When effect sizes are included in published reports, they allow clinicians to develop a sense of the relative strength of the specific instrument [29]. In the present study, only three studies self-reported effect sizes, and only two provided the raw data necessary for precise calculation. Not only would this data provide more conclusive information regarding the treatment effectiveness of SFA, it would also allow for the examination across subgroups (e.g. fluent versus non-fluent, acute versus non-acute). These results would allow us to definitively determine whether SFA may be more effective for individuals with fluent or non-fluent aphasia or for individuals with mild versus moderate aphasia. It is critical that future studies report effect sizes in order to allow for future metaanalysis.

Disclosure of interest

The authors declare that they have no conflicts of interest concerning this article.

2. Version française

2.1. Introduction

L'anomie est le symptôme le plus fréquent et le plus persistant de l'aphasie [1,2]. L'anomie survient secondairement à un événement neurologique (AVC, lésion cérébrale...) et inhibe la capacité de formulation de langage, même au niveau des mots les plus élémentaires. L'anomie est une caractéristique qui définit l'aphasie; elle recouvre de multiples sous-types de celle-ci et concerne toutes les formes grammaticales des mots (noms, verbes...) [3,4]. Les déficits lexicaux nominatifs déteignent négativement sur la capacité des personnes atteintes de communiquer leurs désirs et besoins et de s'engager et s'intégrer à des activités sociales. Certaines études indiquent que des personnes avec aphasie (PAA) ont tendance à être plus dérangées par leur handicap en expression orale que par leurs handicaps en matière de lecture, d'écriture et de compréhension orale. Qui plus est, leurs difficultés d'expression retentissent sur la façon dont les PAA sont vues par les autres dans la vie de tous les jours [2,4]. L'identification d'un traitement efficace susceptible de résorber les déficits nominatifs chez les PAA revêt une importance critique.

Il existe déjà un certain nombre d'approches au traitement centrées sur l'amélioration des capacités linguistiques expressives chez les PAA atteintes d'anomie. Elles partent de l'idée que de manière typique, l'anomie est le résultat de handicaps sémantiques globaux et qu'elle est le reflet d'une implication insuffisante des aspects critiques qui distinguent les concepts les uns des autres [5]. De manière générale, dans les traitements ciblant l'anomie les approches sémantiques sont préférées à des approches phonologiques. Parmi les exemples figurent la nomination induite par des circonlocutions [6], le repérage personnalisé (*personalized cueing*) [7–9] et l'analyse des traits sémantiques (ATS) [10–13]. Dans ce recensement, c'est en mettant l'accent sur l'efficacité clinque que nous allons aborder le traitement de l'anomie chez des adultes souffrant d'atteintes neurologiques.

L'ATS fut initialement présentée par Ylvisaker et Szekeres [14]; par la suite, elle a été développée par Massaro et Tompkins [15]. Il s'agit d'un traitement assez répandu visant en cas d'anomie à améliorer les capacités langagières nominatives et expressives en proposant une méthode organisée d'activation des réseaux sémantiques [16]. L'analyse emploie une technique de repérage dans laquelle on demande aux PAA de produire des mots sémantiquement associés au mot cible qu'ils n'arrivent pas à se rappeler [5]. Par exemple, supposons que le mot cible est « tasse ». Les tentatives de repérage (cues) pourraient comporter des questions ayant trait à son utilisation (Qu'est-ce que vous faites avec ?), à ses propriétés (À quoi cet objet ressemble-t-il ?), aux lieux possibles de son utilisation (À quel endroit le trouvez-vous ?), à la catégorie dont il relève et aux

associations possibles (À quels autres objets est-il similaire ?). Puisqu'il a été proposé que l'anomie résulte d'un réseau sémantique entravé, le but de la thérapie consiste à modifier la connectivité du réseau sémantique en développent de nouveau le réseau endommagé [5]. De manière hypothétique, l'ATS améliorerait la récupération d'informations conceptuelles en accédant aux réseaux sémantiques qu'elle développerait [15]. L'activation renforcée du réseau sémantique entourant le mot cible augmenterait les chances de récupération du mot et pourrait contribuer à la réparation du système sémantique endommagé [5,12,15].

Une revue récente de Boyle [17] a évalué l'efficacité de l'ATS. Cet auteur a recensé sept études différentes, qui ont toutes rapporté des résultats positifs soulignant l'efficacité de l'ATS en atténuant l'anomie chez les personnes avec des handicaps neurologiques. Cela dit, l'importance ou l'ampleur de l'effet traitement n'est considérée que dans trois des sept études [18-20]. Une limitation significative de la revue de Boyle [17] concerne l'absence de toute tentative de calculer l'ampleur de l'effet en utilisant les données rapportées dans les études incluses. À cause de l'absence de ce calcul dans les autres études publiées, on peut difficilement conclure à l'efficacité du traitement. Par conséquent, nous avons établi deux objectifs. Premièrement, nous allons élargir la revue de Boyle [17] en évoquant des recherches nouvelles. Deuxièmement, nous allons appliquer des méthodes statistiques en vue de déterminer l'ampleur de l'effet des traitements répertoriés dans les études incluses et de répondre de manière satisfaisante à la question clinique suivante: Dans des patients souffrant d'aphasie non dégénérative, est-ce que l'analyse des traits sémantiques améliore les capacités de désignation par rapport au mot cible?

2.2. Méthodes

Une revue de la littérature a été effectuée afin d'identifier des études enquêtant sur l'ATS comme intervention primaire en cas d'anomie chez les PAA. Jusque dans leurs versions de juin 2013, sept bases de données électroniques étaient explorées : Academic Search Premier, AgeLine, CINAHL, ERIC, Medline, PsycInfo et Linguistics and Language Behavior Abstracts. Une revue additionnelle a eu lieu en passant au peigne fin les journaux de l'American Speech-Language and Hearing Association, et les références à tous les articles pertinents étaient également épluchées afin d'identifier encore d'autres études utilisables. Parmi les mots-clés retenus figuraient : aphasie, analyse des traits sémantiques, désordres, trouble du langage, repérages sémantiques, anomie, traitement du langage et traitement nominatif.

Lors de la revue 88 articles ont été identifiés, mais dès que le premier auteur a passé en revue les titres et les résumés, 70 articles ont été éliminés. Une étude a été exclue de la revue :

- si elle n'était pas expérimentale ;
- si elle n'abordait pas l'ATS comme traitement ;
- si elle ne concernait pas des adultes souffrant d'atteintes neurologiques.

Par contre, dans un premier temps les études étaient retenues :

- si elles étaient rédigées en anglais ;
- si elles avaient fait l'objet d'une publication entre 1980 et juin 2013 dans un journal avec comité de lecture.

Ensuite, le premier auteur a passé en revue les 18 articles restants dont 7 étaient exclus parce qu'ils associaient l'ATS à d'autres méthodes de traitement (thérapie de groupe ou thérapie de stimulation sémantique) ou parce qu'ils employaient une approche hybride à l'ATS. Afin de rester en conformité avec l'objectif de la revue, seules les études abordant l'ATS telle qu'elle avait été initialement conçue par Massaro et Tompkins [15] ont été incluses de manière définitive. Le processus par lequel les articles à inclure ont été identifiés dans le cadre de notre revue systématique fondée sur des preuves est résumé sur la Fig. 1.

Deux orthophonistes certifiés et habilités ont évalué de manière critique les études sélectionnées en termes de qualité méthodologique. Dix ont été évalués en utilisant l'échelle SCED (Single Case Experimental Design) [21]. N'étant pas une étude avec plan de recherche centré sur sujet unique, l'échelle SCED n'était pas appliquée au travail de Marcotte et Ansaldo [22]. Le SCED est une échelle de 11 utilisée dans l'évaluation d'études expérimentales focalisées sur un seul cas. Chaque évaluateur indépendant accordait soit une note de 1, si l'étude abordait de manière satisfaisante tel ou tel élément spécifique, soit une note de 0, si l'élément était abordé de manière insatisfaisante ou pas du tout. Suite aux notations indépendantes, la fiabilité inter-évaluateurs a été calculée selon les statistiques kappa pondérées de Cohen [23]. Le coefficient kappa pondéré est la proportion de concordance ne relevant pas du hasard et qui tient compte du degré de discordance des deux évaluateurs indépendants. Dans notre enquête, le score kappa pondéré s'élevait à 0,656, ce qui indique une bonne fiabilité inter-évaluateurs [24]. En cas de désaccord, un score moyen a été calculé. Le premier auteur a choisi de façon aléatoire trois études et a recalculé les scores SCED en vue de déterminer le degré de fiabilité intra-évaluateurs. Afin de réduire les biais éventuels et de s'assurer de la non-dépendance des évaluations l'une de l'autre, le second calcul des scores a été complété deux semaines après le premier [25]. Le score kappa pondéré mesurant la fiabilité intra-évaluateurs s'élevait à 1,0, ce qui indique une parfaite fiabilité intra-évaluateurs.

On sait que l'efficacité clinique est une variable essentielle à prendre en considération dans n'importe quel traitement. En même temps, il importe d'interpréter l'efficacité clinique dans le contexte de telle ou telle étude. Afin de renforcer l'interprétation des résultats de notre revue systématique, des niveaux de preuve prenant comme modèle la hiérarchie [26] de niveaux de preuve de l'ASHA (American Speech and Hearing Association) ont été déterminés. Cette hiérarchie a été choisie à cause de son applicabilité à l'efficacité d'interventions auprès de personnes en proie à des troubles de la communication. Deux évaluateurs indépendants ont examiné chaque étude et lui ont attribué tel ou tel niveau de preuve. Suite à leurs notations

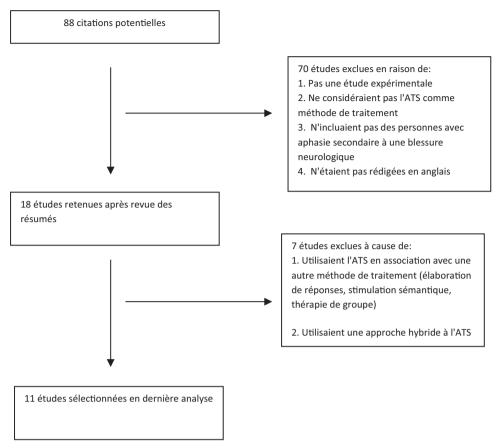


Fig. 1. Processus d'identification des études à inclure.

indépendantes, le kappa pondéré a été calculé afin d'évaluer la fiabilité des attributions de niveaux de preuve. Le premier auteur a choisi 30 % des études de façon aléatoire pour les besoins de la réévaluation et afin de calculer la fiabilité intraévaluateurs. La fiabilité inter-évaluateurs et la fiabilité intraévaluateurs étaient toutes les deux excellentes (k = 1,0).

Dans le but de déterminer l'efficacité de l'ATS comme méthode de traitement de l'aphasie, les ampleurs de l'effet étaient calculées dans les études dont les données étaient suffisamment importantes. Mais avant de procéder aux calculs, il état nécessaire de déterminer les valeurs individuelles de chaque ensemble d'éléments avant et après le traitement. Par conséquent, l'ampleur de l'effet était calculée lorsque les données des réponses étaient présentées dans des incréments de nombres entiers en utilisant une variation de la statistique d de Cohen [27] telle qu'elle a été décrite par Busk et Serlin [28]. La formule suivante a été utilisée :

$$d_1 = \frac{xA_2 - xA_1}{sA_1}$$

où A_2 et xA_1 désignent respectivement les phases post-traitement et pré-traitement, alors que xA est la moyenne des données recueillies pendant telle et telle phase, et sA_1 est l'écart-type correspondant. Bien que les auteurs préconisent trois périodes d'observation pendant la phase post-traitement en vue d'aboutir à une estimation maximalement fiable de l'importance de

l'effet, il est mathématiquement possible de calculer les ampleurs de l'effet à la suite d'une seule période d'observation rapportée [29]. Dans l'optique de notre étude, les ampleurs de l'effet ont été calculées dès qu'au moins une période d'observation a été rapportée. Le degré d'importance du niveau de performance a été déterminé selon les points de repère dans les études de récupération, références déjà décrites par Beeson et Robey [29]. De tels *benchmarks* ont été fixés à 4,0, 7,0 et 10,1 indiquant des ampleurs de l'effet respectivement petites, moyennes et grandes.

Lorsqu'il n'était pas possible de calculer la statistique d, le pourcentage de données mutuellement exclusives (PDME) a été calculé. Il s'agit de la méthode actuellement la plus répandue de calculer l'ampleur de l'effet dans des plans de recherche centrés sur un sujet unique [30,31]. PDME est le pourcentage des points de données de la phase B (la phase de traitement) qui ne recoupent pas (dans un sens favorable) les points de données de la phase A (éléments de référence ou pas de traitement). Afin de déterminer l'importance de l'effet, nous avons utilisé les points de repère proposés par Scruggs et al. [32]. Un pourcentage PDME supérieur à 90 % dénote un traitement hautement efficace, alors qu'un PDME de 70 à 90 % reflète des résultats moyens et un PDME de 50 à 70 % traduit des effets discutables. Tout PDME inférieur à 50 % indique une intervention inefficace puisque celle-ci n'aura pas donné lieu à une amélioration du comportement du sujet.

2.3. Résultats

Les dix études incluses dans la revue comportaient toutes des plans de recherche centrés sur un sujet unique. L'une des études comprenait les phases pré-traitement et post-traitement. Vingt-quatre participants figuraient dans la revue. Les variables démographiques intéressant l'ensemble des participants sont présentées dans le Tableau 1. Il existait une hétérogénéité significative des participants en termes d'âge et de laps de temps après le début de la maladie. L'âge des participants s'échelonnait de 24 à 85 ans, 1'âge moyen étant de 60,75 ans (ET = 14,29), et le laps de temps après le début de la maladie s'échelonnait de 4 à 187 mois, le laps moyen étant de 53,74 mois (ET = 50,44). Quinze hommes (63 %) et 9 femmes (37 %) étaient représentés. L'aphasie non fluente constituait le sous-type le plus représenté dans la revue (n = 17; 71 %), alors que les sept autres participants (29 %) étaient atteints d'aphasie fluente. Vingt et un participants avaient été victimes d'un accident cardiovasculaire, alors que les trois autres participants présentaient avec aphasie suite à un traumatisme crânien. La gravité de l'aphasie n'était indiquée que dans cinq des 11 études incluses dans la revue.

Les scores SCED s'échelonnaient de 8,0 à 10,5, le score moyen étant de 9,3 (score maximum : 11,0) (Tableau 2). En

fonction des scores obtenus, les études étaient distribuées en termes de niveaux de preuve [26]. Dix études étaient considérées comme bien conçues et quasi-expérimentales ; elles recevaient la notation correspondant au niveau IIb. Quant à l'étude de Marcotte et Ansalso [22], elle entrait en ligne de compte comme étude observationnelle contrôlée et recevait la notation correspondant au niveau III. La prévalence de scores SCED élevés et de niveaux de preuve IIb traduisent l'excellente qualité méthodologique et la grande rigueur caractérisant les études passées en revue.

Trois études ont comporté des auto-déclarations sur l'ampleur de l'effet chez les participants [3,18,19]. Le calcul des ampleurs se rapportait à cinq participants (Tableau 3). Dès que les données sur deux essais au minimum de la variable dépendante avaient été recueillies et rapportées, une ampleur de l'effet moyenne a été calculée [18,20,33]. Une grande ampleur de l'effet a été constatée chez deux participants (d = 18,76; d = 10,56). Un effet modéré a été relevé chez six participants (d = 7,95; d = 9,28; d = 7,48; d = 7,79; d = 7,32; d = 9,617), et un effet faible chez huit autres (d = 1,79; d = 3,86; d = 5,54; d = 2,97; d = 1,26; d = 5,01, d = 5,32; d = 6,35). Une prévalence élevée de faibles ampleurs indique que le traitement n'était pas aussi efficace dans l'atténuation de l'anomie auprès les huit derniers qu'auprès des huit premiers participants. Par

Tableau 1
Caractéristiques démographiques des participants

Étude	n	Participant	Âge (années)	Sexe	Étiologie	TAD (mois)	Type d'aphasie	Sévérité d'aphasie
Boyle et Coelho, 1995 [12]	1	P1	57	M	ACVG	65	Non fluent	NR
Boyle, 2004 [11]	2	P1 P2	70 80	M M	ACVG ACVG	15 14	Fluent Fluent	NR NR
Coelho et al., 2000 [13]	1	P1	52	M	TC	17	Fluent	Modéré- sévère
Davis et Stanton, 2005 [34]	1	P1	59	F	ACV	4	Fluent	NR
Hashimoto, 2011 [33]	1	P1	72	F	ACV	NR	Non fluent	Modéré- sévère
Marcotte et Ansaldo, 2010 [22]	1	P1	66	M	ACV	84	Non fluent	Sévère
Massaro et Tompkins, 1994 [15]	2	P1 P2	24 28	M F	TC TC	60 144	Non fluent Non fluent	NR NR
Peach et Reuter, 2010 [19]	2	P1 P2	77 62	F F	ACVG ACVG	4 14	Fluent Fluent	Léger Modéré
Rider et al., 2008 [18]	3	P1 P2 P3	73 55 62	M F M	ACVG ACVG ACVG	26 45 126	Non fluent Non fluent Non fluent	NR NR NR
Fambaugh et Ferguson, 2007 [3]	1	P1	74	F	ACVG	50	Non fluent	Modéré
Wambaugh et al., 2013 [20]	9	P1 P2 P3 P4 P5 P6 P7 P8	85 59 61 47 59 52 66 64	M F M M M M M	ACV ACV ACV ACV ACV ACV ACV ACV	126 42 31 187 65 13 65 18	Non fluent	NR NR NR NR NR NR NR NR

NR: non rapporté; ACV: accident cérébrovasculaire; G: hémisphère gauche; TAD: laps de temps après début de la maladie.

Tableau 2 Évaluation critique de la littérature.

Étude	Méthodologie	Score qualité SCED	
Boyle et Coelho, 1995 [12]	A-B sujet unique	8	
Boyle, 2004 [11]	Niveaux de base parallèles avec observations comportementales	10	
Coelho et al., 2000 [13]	A-B sujet unique	8,5	
Davis et Stanton, 2005 [35]	Niveaux de base parallèles avec observations comportementales	8	
Hashimoto, 2011 [33]	Niveaux de base parallèles avec observations comportementales	8	
Marcotte et Ansaldo, 2010 [22]	Étude pré-/post-traitement	N/A	
Massaro et Tompkins, 1994 [15]	Niveaux de base parallèles avec observations comportementales	10	
Peach et Reuter, 2010 [19]	Série chronologique avec observations comportementales sur cas unique	10	
Rider et al., 2008 [18]	Multiples observations comportementales	10,5	
Wambaugh et Ferguson, 2007 [3]	Niveaux de base parallèles avec observations comportementales	10	
Wambaugh et al., 2013 [20]	Niveaux de base parallèles avec observations comportementales	10	

Tableau 3 Ampleurs de l'effet et pourcentages de données mutuellement exclusives (PDME) dans les études incluses.

Participant	Type d'aphasie	d de Cohen	PDME	Importance de l'effet
P1	Non fluent		100 %	Hautement efficace
P1 P2	Non fluent Fluent	18,48 ^b CND	100 %	Grand effet Hautement efficace
P1	Fluent	CND	100 %	Hautement efficace
P1	Fluent	CND	91,67 %	Hautement efficace
P1	Non fluent	10,56 ^a		Grand effet
P1	Non fluent	CND	CND	N/A
P1 P2	Non fluent Non fluent	CND CND	100 % 100 %	Hautement efficace Hautement efficace
P1 P2	Fluent Fluent	1,79 CND	85 %	Effet faible Moyennement efficace
P1 P2 P3	Non fluent Non fluent Non fluent	3,86 ^a 5.54 ^a 2,97 ^a		Effet faible Effet faible Effet faible
P1	Non fluent	6,35 ^b		Effet moyen
P1 P2 P3 P4 P5 P6 P7 P8	Non fluent	5,32 ^a 7,95 ^a 9,28 ^a 7,48 1,26 ^a 5,01 ^a 7,79 ^a 7,32 ^a		Effet faible Effet moyen Effet moyen Effet moyen Effet faible Effet faible Effet moyen Effet moyen Effet moyen Effet moyen
	P1 P1 P2 P1 P1 P1 P1 P1 P1 P1 P2 P1 P2 P1 P2 P1 P2 P3 P1 P2 P3 P4 P5 P6 P7	P1 Non fluent P1 Non fluent P2 Fluent P1 Fluent P1 Fluent P1 Non fluent P1 Non fluent P1 Non fluent P2 Non fluent P1 Fluent P2 Fluent P1 Non fluent P2 Non fluent P3 Non fluent P1 Non fluent P2 Non fluent P3 Non fluent P4 Non fluent P5 Non fluent P6 Non fluent P7 Non fluent P8 Non fluent	P1 Non fluent P1 Non fluent P2 Fluent P1 Fluent P1 Fluent P1 Fluent P1 Non fluent P1 Non fluent P1 Non fluent P2 Non fluent P3 Non fluent P4 Non fluent P5 Non fluent P6 Non fluent P7 Non fluent P7 Non fluent P7 Non fluent P8 Non fluent P7 Non fluent P8 Non fluent P8 Non fluent P7 Non fluent P8 Non fluent P8 Non fluent P8	P1 Non fluent 18,48b P2 Fluent CND 100 % P1 Fluent CND 100 % P1 Fluent CND 91,67 % P1 Non fluent CND 91,67 % P1 Non fluent CND CND P1 Non fluent CND 100 % P2 Non fluent CND 100 % P1 Fluent 1,79 P P2 Fluent CND 85 % P1 Non fluent 3,86a P P2 Non fluent 5.54a P P3 Non fluent 2,97a P P1 Non fluent 5,32a P P2 Non fluent 7,95a P P3 Non fluent 7,95a P P3 Non fluent 7,48 P4 Non fluent 7,48 P5 Non fluent 5,01a P6 Non fluent 7,79a P8 Non fluent 7,32a

PDME : pourcentage de données mutuellement exclusives.

ailleurs, chez l'un des participants à l'étude de Marcotte et Ansaldo [22], l'ampleur de l'effet et le PDME n'ont pas pu être calculés.

Chez sept participants, le pourcentage PDME a été calculé. Une grande ampleur de l'effet (PDME > 90 %) a été constatée dans six cas et un effet modéré dans l'autre cas (PDME = 85 %). Lorsqu'on évalue l'efficacité clinique à l'aide du PDME, chez la majorité des patients le traitement s'est avéré hautement efficace. Aucun participant n'a enregistré un score PDME traduisant l'inefficacité du traitement.

2.4. Discussion

La lecture des résultats indique que l'ATS constitue une intervention efficace dans l'amélioration de la désignation par rapport au mot cible d'éléments abordés dans la thérapie chez des personnes avec aphasie non dégénérative. D'une part, il y avait des résultats positifs auto-déclarés ; d'autre part, les calculs d'ampleur de l'effet mettent en évidence l'efficacité modérée ou élevée du traitement chez la majorité des patients considérés dans la revue. Par ailleurs, même si le traitement

^a Calcul moyen.

^b Calculé par les auteurs de ce présent papier, à partir des données présentées dans les articles.

était particulièrement efficace en améliorant la désignation par rapport au mot cible des éléments abordés (*trained items*), dans un nombre appréciable des études incluses, la désignation était généralisable au point de s'étendre à des éléments non abordés (*untrained items*), voire à des discours associés. Ce type de résultat mérite d'être approfondi en vue de faire le point sur les facteurs (intensité, dosage...) qui pourraient favoriser la généralisation.

Notre revue n'est pas pourtant exempte de limitations dont la discussion détaillée va suivre.

2.4.1. Variations en termes de participants à l'étude

Pour les besoins de l'étude, parmi les personnes incluses figuraient des participants dont la récupération était dans la phase aiguë. Bien que la récupération spontanée pût donner lieu à des résultats de traitement favorables, tous les participants en phase aiguë de récupération avaient présenté préalablement au traitement des bases de référence stables. En outre, les participants en question n'ont pas présenté à l'issue du traitement une ampleur de l'effet ou un pourcentage PDME anormalement grand, d' = 1.79 [19], PDME = 85 % [19] et PDME = 91,67 % [35]. En ce qui concerne le laps de temps entre le début de la maladie et le début du traitement (TAD, Tableau 1), bien que ce laps ait été assez variable, dans 92 % (22/24) des cas le début du traitement avait eu lieu dans une phase post-aiguë (plus de 6 mois après le début de la maladie). Ce constat n'a rien de surprenant dans la mesure où dans la littérature sur l'aphasie, la plupart des études avec des plans de recherche à sujet unique concernent des personnes dont le TAD est significativement élevé; le choix de ce type de patient est effectué en vue d'obtenir une série stable d'observations et de tenir compte de la récupération spontanée [29]. L'inconvénient, c'est que l'importance des effets rapportés est fort susceptible de sous-estimer l'importance potentielle des effets chez des personnes souffrant d'aphasie aiguë, puisque des progrès particulièrement grands sont souvent observés lors de la phase aiguë de la récupération [5]. Qui plus est, dans la littérature sur l'aphasie l'âge et le TAD sont très régulièrement cités comme des indicateurs pronostiques significatifs de récupération de capacités langagières [5,37]. Dans les études encore à venir, des questions devront être posées sur l'efficacité de l'ATS dans des tranches d'âge et des TAD divers et variés.

Le type d'aphasie rencontrée chez les participants à l'étude était varié : anomie, aphasie motrice transcorticale, aphasie de Broca. En termes des grandes catégories d'aphasie fluente et non fluente, celle-ci a prédominé (n=17; 71%), alors que les sept autres participants (29%) étaient atteints d'aphasie fluente (Tableau 1). Les résultats semblent indiquer qu'en tant que traitement de l'anomie, l'ATS serait plus efficace chez des aphasiques fluents que chez des aphasiques non fluents. Bien que des difficultés dans la recherche et repérage de mots puissent se présenter tant dans les aphasies fluentes que dans les aphasies non fluentes, la tâche consistant à décrire des aspects pertinents d'un mot cible risque d'être plus compliquée chez les aphasiques non fluents. Les recherches précédentes suggèrent que les aphasiques fluents présentent des déficits significatifs en connaissance des catégories [38,39], ce qui rendrait plus

efficace auprès de ce groupe des traitements ciblant de tels déficits. Cela dit, seules des recherches encore à venir assiéront davantage des comparaisons des bénéfices respectifs de l'ATS chez des personnes fluentes et non fluentes. Par conséquent, les cliniciens s'intéressant à l'ATS comme traitement doivent tenir compte du type d'aphasie.

2.4.2. Variations en termes de l'intervention

Tant les cliniciens que les chercheurs s'aviseront du fait que l'ATS n'est pas toujours effectuée selon une procédure standardisée. Si certaines études ont fait état de périodes de traitement relativement longues, d'autres études ont rendu compte de séances de traitement plus fréquentes en un laps de temps relativement court. Dans l'étude de Rider et al. [18], le traitement était composé de deux ou trois séances hebdomadaires d'une heure pendant une période de cinq semaines. Quant à Davis et Stanton [34], ils ont appliqué un protocole de traitement similaire consistant en deux séances thérapeutiques hebdomadaires pendant six semaines. Seules des recherches supplémentaires permettront de déterminer le dosage approprié et la durée du traitement le plus apte à optimiser l'efficacité clinique. En outre, pendant que certaines études ont ciblé des exemples atypiques (« aigrette »), d'autres études ont ciblé des exemples plus typiques (« robin »). L'exécution standardisée de l'ATS selon l'objectif initial du traitement et par rapport aux stimuli et aux dosages permettraient la réalisation de méta-analyses dont le but consisterait à déterminer de manière plus concluante l'efficacité relative de tel ou tel traitement.

2.4.3. Variations en termes de mesures de résultats

Il existe une hétérogénéité significative dans les mesures de résultats utilisées dans les études incluses. Par exemple, Peach et Reuter [19] ont examiné l'effet de l'ATS sur la production du discours, c'est-à-dire la productivité verbale (# de mots par Tunit ou unité phrastique minimale), la valeur informative (unités de contenu informatif ; %UCI) et la désignation d'éléments non abordés. Puisque de telles mesures n'étaient pas directement associées aux résultats du traitement, il est difficile de calculer des ampleurs d'effet significatives. Boyle et Coelho [12], Davis et Stanton [34] et Boyle [11] ont tous analysé le pourcentage de dessins désignés correctement comme première variable dépendante de leurs études, mais ils ont également calculé en termes de %UCI. En complément à ces mesures, Boyle et Coelho [12] ont utilisé le Communicative Effectiveness Index [39] en vue de juger de la validité sociale des changements intervenus. Et en complément au pourcentage de dessins désignés correctement, Rider et al. [18] ont examiné la diversité lexicale (d) et la performance dans des évaluations langagières standardisées, dont la Western Aphasia Battery [40], le Boston Naming Test [41] et la Reading Comprehension Battery for Aphasia [42].

2.4.4. Limitations de l'analyse statistique

L'utilisation d'un nombre relativement élevé et diversifié de mesures de résultats a compliqué toute tentative de déterminer de manière définitive l'impact clinique du traitement. Pour cette raison-là, le recours à la méta-analyse s'impose, et elle présente plusieurs avantages [18]. Premièrement, elle rend possible la génération d'un effet global en combinant des études et en établissant l'importance de l'effet traitement. Deuxièmement, si l'effet dans les différentes études est uniforme, elle le précise ; si l'effet n'est pas uniforme, elle représente un mécanisme apte à étudier l'hétérogénéité des participants. Troisièmement, la méta-analyse permet des comparaisons concernant des sous-groupes (sexe, âge, type d'aphasie, sévérité de l'aphasie).

Une revue récente des plans de recherche des études de sujets uniques démontre que le PDME est le paramètre des résultats le plus fréquemment utilisé dans des revues systématiques [30]. Cela dit, son utilisation ne va pas sans limitations. Il est important de faire remarquer qu'il n'existe pas de référence ou benchmark du PDME dans les études publiées sur l'aphasie, ce qui rend problématique l'estimation de l'importance de l'effet de tel ou tel traitement de celle-ci ; l'interprétation des résultats des études doit être effectuée prudemment. En outre, nous avons noté que comparativement à l'ampleur des effets des calculs avec d, le PDME risque de surestimer des effets traitement. Par exemple, Rider et al. [18] ont rapporté chez trois participants de petites ampleurs de l'effet. Or suite au calcul du PDME, nous sommes arrivés à des scores hautement efficace (PDME = 90 %), modérément efficace (PDME = 80 %) et discutablement efficace (PDME = 50 %). De manière similaire, Peach et Reuter [19] ont calculé chez l'un des participants une petite ampleur de l'effet (d = 1,79) se rapportant à la productivité verbale pendant la phase de traitement. Mais là encore, le calcul du PDME a donné lieu à un score modérément efficace (PDME = 78 %). De telles comparaisons suggèrent que le PDME constitue une méthode plus « ouverte » d'analyse de l'efficacité des traitements. Il importe à ce propos de se rappeler que le PDME est le pourcentage des points de données de la phase B (phase de traitement) qui ne recoupent pas (dans un sens favorable) des points de données de la phase A (éléments de référence ou pas de traitement). Une personne pourrait améliorer de 10 % sa capacité de désignation par rapport au mot cible, alors qu'une autre personne pourrait améliorer de 50 % sa capacité de désignation, mais au cas où il n'y aurait pas de recoupement des points de données respectifs de la phase A et de la phase B, les deux patients remporteraient le même score PDME! D'une part, le calcul en question permet de décider de l'existence d'effets traitement ; d'autre part, il n'examine pas de manière complète l'importance desdits effets. Seules des recherches encore à venir pourraient permettre de déterminer la corrélation entre les scores PDME et la statistique en d, et donc de déterminer le degré de fiabilité du PDME comme mesure de l'importance des effets de tel ou tel traitement de l'aphasie.

Afin de mener à bien une vraie méta-analyse, des ampleurs de l'effet doivent être établies. Puisque de telles ampleurs sont quantifiées comme unités d'écarts-types, d'une étude à l'autre elles peuvent être comparées et combinées dans des méta-analyses [29]. Bien que ces ampleurs soient désormais fréquemment indiquées dans des études de groupes, elles continuent à être généralement absentes d'études avec plan de

recherche centré sur un sujet unique. Lorsqu'elles sont incluses dans des rapports publiés, elles permettent aux cliniciens de se faire une idée de sa force relative de tel ou tel outil [29]. Dans notre étude, seules trois études ont rapporté les ampleurs de l'effet, et deux seulement ont fourni des données brutes nécessaires à des calculs précis. Or tout en fournissant des informations plus concluantes sur l'efficacité de l'ATS, de telles données permettraient également l'analyse approfondie des sous-groupes (fluents et non fluents, aigus et non aigus). Les résultats nous permettraient de décider : Est-ce que l'ATS est efficace chez des personnes avec l'aphasie fluente ou avec l'aphasie non fluente ? Avec de l'aphasie modérée ou de l'aphasie légère ? Il sera d'une importance critique que les études à venir précisent les ampleurs de l'effet, ce qui faciliterait la méta-analyse.

Déclaration d'intérêts

Les auteurs déclarent ne pas avoir de conflits d'intérêts en relation avec cet article.

References

- [1] Rose M, Douglas J. Treating a semantic word production deficit in aphasia with verbal and gesture methods. Aphasiology 2008;22:20–41.
- [2] Nickels L. Therapy for naming disorders: revisiting, revising, and reviewing. Aphasiology 2002;16:935–79.
- [3] Wambaugh JL, Ferguson M. Application of semantic feature analysis to retrieval of action names in aphasia. J Rehabil Res Dev 2007;44:381–94.
- [4] Brookshire R. Introduction to neurogenic communication disorders. 6th ed, St. Louis, MO: Mosby, Inc; 2003.
- [5] Chapey R. Language intervention strategies in aphasia and related neurogenic communication disorders. 5th ed, Baltimore, MD: Lippincott Williams & Wilkins; 2010.
- [6] Francis DR, Clark N, Humphreys GW. Circumlocution-induced naming: a treatment for effecting generalization in anomia? Aphasiology 2002; 16:243–60.
- [7] Freed DB, Celery K, Marshall RC. Effectiveness of personalized and phonological cueing on long-term naming performance by aphasic subjects: a clinical investigation. Aphasiology 2004;18:743–57.
- [8] Freed DB, Marshall RC. The effects of personalized cueing on the longterm naming of realistic visual stimuli. Am J Speech Lang Pathol 1995;4:105–8.
- [9] Marshall RC, Freed DB. The personalized cueing method: from the laboratory to the clinic. J Neurol Phys Ther 2006;15:103–11.
- [10] Boyle M. Semantic feature analysis treatment for aphasic word retrieval impairments: what's in a name? Top Stroke Rehabil 2010;17:411–22.
- [11] Boyle M. Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. Am J Speech Lang Pathol 2004;13:236–49.
- [12] Boyle M, Coelho C. Application of semantic feature analysis as a treatment for aphasic dysnomia. Am J Speech Lang Pathol 1995;13:236–49.
- [13] Coelho C, McHugh R, Boyle M. Semantic feature analysis as a treatment for aphasic dysnomia: a replication. Aphasiology 2000;14:133–42.
- [14] Ylvisaker M, Szekeres S.;1; Cognitive-language intervention with braininjured adolescents and adults. Annual Convention of the Illinois Speech-Language-Hearing Association; November 1985.
- [15] Massaro M, Tompkins C. Feature analysis for treatment of communication disorders in traumatically brain-injured patients: an efficacy study. Clin Aphasiol 1994;22:245–56.
- [16] Coelho C, Sinotte M. J.D. Schuell's stimulation approach to rehabilitation. In: Chapey R, editor. Language intervention strategies in aphasia and related neurogenic communication disorders. 5th ed, Baltimore: Lippincott Williams & Wilkins; 2008.

- [17] Boyle M. Semantic feature analysis treatment for aphasic word retrieval impairments; what's in a name? Top Stroke Rehabil 2010;17:411–22.
- [18] Rider JD, Wright HH, Marshall RC, Page JL. Using semantic feature analysis to improve contextual discourse in adults with aphasia. Am J Speech Lang Pathol 2008;17:161–72.
- [19] Peach R, Reuter K. A discourse-based approach to semantic feature analysis for the treatment of aphasic word retrieval failures. Aphasiology 2010;24:971–90.
- [20] Wambaugh JL, Mauszycki S, Cameron R, Wright S, Nessler C. Semantic feature analysis: incorporating typicality treatment and mediating strategy training to promote generalization. Am J Speech Lang Pathol 2013;22. S334–S69.
- [21] Tate R, McDonald K, Perdices M, Tougher L, Schultz R, Savage S. Rating the methodological quality of single-subject designs and n-of-1 trials: introducing the Single-Case Experimental Design (SCED) Scale. Neuropsychol Rehabil 2008;18:385–401.
- [22] Marcotte K, Ansaldo AI. The neural correlates of semantic feature analysis in chronic aphasia: discordant patterns according to the etiology. Semin Speech Lang 2010;31:52–63.
- [23] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46.
- [24] Altman D. Practical statistics for medical research. London: Chapman and Hall; 1991.
- [25] Caplan D. Language: structure, processing, and disorders. Cambridge, Mass: MIT Press; 1992.
- [26] Lock S, Armstrong L. Cohesion analysis of the expository discourse of normal, fluent aphasic and demented adults: a role in differential diagnosis? Clin Linguist Phon 1997;11:299–317.
- [27] Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed, Hillsdale, NJ: Erlbaum: 1988.
- [28] Busk PL, Serlin RC. Meta-analysis for single-case research. In: Kratoch-will TR, Levin JR, editors. Single-case research design and analysis. Hillsdale, NJ: Lawrence Erlbaum; 1992. 1887–212.
- [29] Beeson P, Robey R. Evaluating single-subject treatment research: lessons learned from the aphasia literature. Neuropsychol Rev 2006;16:161–9.

- [30] Schlosser RW, Lee DL, Wendt O. Application of the percentage of nonoverlapping data (PND) in systematic reviews and meta-analyses: a systematic review of reporting characteristics. Evid Based Commun Assess Intervent 2008;2:163–87.
- [31] Gast D. Single subject research methodology in behavioral sciences. New York: Routledge; 2010.
- [32] Scruggs T, Mastropieri M, Casto G. The quantitative synthesis of singlesubject research: methodology and validation. Rem Spec Educ 1987; 8:24–33.
- [33] Hashimoto N. The use of a modified semantic features analysis approach in aphasia. J Commun Disord 2011;44:459–69.
- [34] Davis L, Stanton S. Semantic feature analysis as a functional therapy tool. Contemp Issues Commun Sci Disord 2005;32:85–92.
- [35] Haynes WO, Pindzola RH. Diagnosis and evaluation in speech pathology. 6th ed, Boston: Pearson Education, Inc; 2004.
- [36] Shelton JR, Caramazza A. Deficits in lexical and semantic processing: implications for models for normal language. Psychon Bull Rev 1996:6:5–27.
- [37] Kiran S, Thompson C. The role of semantic complexity in treatment of naming deficits: training semantic categories in fluent aphasia by controlling exemplar typicality. J Speech Lang Hear Res 2003;46:773–87.
- [38] Nicholas L, Brookshire R. A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. J Speech Hear Res 1993;36:338–50.
- [39] Lomas J, Pickard L, Bester S, Elbard H, Finlayson A, Zoghaib C. The Communicative Effectiveness Index: development and psychometric evaluation of a functional communication measure for adult aphasi. J Speech Hear Disord 1989;54:113–24.
- [40] Kertesz A. Western Aphasia Battery. New York: Grune and Stratton; 1982
- [41] Goodglass H, Kaplan E. The Boston Naming Test. 2nd ed, Hagerstown, MD: Lippincott Williams & Wilkins; 2001.
- [42] LaPointe L, Horner J. Reading Comprehension Battery for Aphasia. 2nd ed, Austin, TX: Brown University Press; 1998.
- [43] Borenstein M. Introduction to meta-analysis. Chichester, UK: John Wiley & Sons; 2009 [xxviii, 421 pp.].